

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 10:23:22 ; Search time 41 Seconds  
(without alignments)

4840.060 Million cell updates/sec

Title: US-09-701-271A-2

Perfect score: 3848

Sequence: 1 MNTKLPKIIISGLFVATAFO.....ELLFITPRIMGFAGNSLRY 769

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organella:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp Vertebrate:\*

14: sp Unclassified:\*

15: sp\_rvitus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID        | Description        |
|------------|-------|-------------|--------|-----------|--------------------|
| 1          | 3666  | 95.3        | 761    | 16 Q9JVM4 | Q9JVM4 neisseria m |
| 2          | 3094  | 80.4        | 766    | 2 Q8ZHF3  | Q8ZHF3 neisseria m |
| 3          | 2988  | 77.7        | 720    | 2 Q80972  | Q80972 neisseria g |
| 4          | 1091  | 28.4        | 714    | 16 Q8XV60 | Q8XV60 ralstonia s |
| 5          | 1050  | 27.3        | 723    | 2 Q8AM59  | Q8AM59 acinetobact |
| 6          | 944.5 | 24.5        | 684    | 16 Q8EK21 | Q8EK21 shewanella  |
| 7          | 911.5 | 23.7        | 633    | 16 Q8PH78 | Q8PH78 xanthomonas |
| 8          | 910   | 23.6        | 648    | 16 Q8PSV7 | Q8PSV7 xanthomonas |
| 9          | 907   | 23.6        | 637    | 16 Q8PGC9 | Q8PGC9 xylella fas |
| 10         | 809.5 | 21.0        | 578    | 16 Q8KXV0 | Q8KXV0 vibrio chol |
| 11         | 800.5 | 20.8        | 589    | 16 Q8DCM2 | Q8DCM2 vibrio vuln |
| 12         | 776   | 20.2        | 901    | 2 Q8ZFG1  | Q8ZFG1 myxococcus  |
| 13         | 574.5 | 14.9        | 412    | 16 Q8Z206 | Q8Z206 salmonella  |
| 14         | 571.5 | 14.9        | 412    | 16 Q8ZLK2 | Q8ZLK2 salmonella  |
| 15         | 570.5 | 14.8        | 412    | 16 Q8CVW9 | Q8CVW9 escherichia |
| 16         | 567.5 | 14.7        | 412    | 16 Q8X818 | Q8X818 escherichia |

|    |       |      |      |           |                     |
|----|-------|------|------|-----------|---------------------|
| 17 | 543   | 14.1 | 374  | 16 Q8ZJF8 | Q8ZJF8 versinia pe  |
| 18 | 473.5 | 12.3 | 444  | 16 Q9CLK3 | Q9CLK3 pasteurella  |
| 19 | 393.5 | 10.2 | 754  | 16 Q8XTG8 | Q8XTG8 ralstonia s  |
| 20 | 380   | 9.9  | 805  | 16 Q8XUS1 | Q8XUS1 ralstonia s  |
| 21 | 363   | 9.4  | 1285 | 16 Q9WUJ3 | Q9WUJ3 thermotoga   |
| 22 | 362.5 | 9.4  | 616  | 2 Q8VRN0  | Q8VRN0 escherichia  |
| 23 | 362.5 | 9.4  | 616  | 2 Q8VPC8  | Q8VPC8 escherichia  |
| 24 | 362   | 9.4  | 698  | 16 Q8XSJ8 | Q8XSJ8 ralstonia s  |
| 25 | 352.5 | 9.2  | 692  | 16 Q8PPJ1 | Q8PPJ1 xanthomonas  |
| 26 | 347   | 9.0  | 500  | 9 Q80264  | Q80264 vibrio chol  |
| 27 | 341.5 | 8.9  | 649  | 2 Q9ZFY0  | Q9ZFY0 pseudomonas  |
| 28 | 338.5 | 8.8  | 708  | 16 Q88A33 | Q88A33 rhizobium l  |
| 29 | 337   | 8.8  | 763  | 16 Q8PGT2 | Q8PGT2 xanthomonas  |
| 30 | 336.5 | 8.7  | 689  | 2 Q8RTI3  | Q8RTI3 pseudodalter |
| 31 | 336   | 8.7  | 705  | 16 Q66850 | Q66850 aquifex aeo  |
| 32 | 333.5 | 8.7  | 775  | 16 Q9PD52 | Q9PD52 xylella fas  |
| 33 | 333   | 8.7  | 554  | 16 Q8CVN5 | Q8CVN5 escherichia  |
| 34 | 329.5 | 8.6  | 591  | 2 Q52291  | Q52291 pseudomonas  |
| 35 | 328.5 | 8.5  | 690  | 16 Q8P5B6 | Q8P5B6 xanthomonas  |
| 36 | 327   | 8.5  | 615  | 16 Q8F3M6 | Q8F3M6 leptospira   |
| 37 | 325.5 | 8.5  | 658  | 2 Q8GBE6  | Q8GBE6 versinia en  |
| 38 | 323.5 | 8.4  | 704  | 16 Q8EKC9 | Q8EKC9 shewanella   |
| 39 | 322.5 | 8.4  | 585  | 2 Q32566  | Q32566 escherichia  |
| 40 | 322.5 | 8.4  | 642  | 2 Q9ZGU0  | Q9ZGU0 escherichia  |
| 41 | 319.5 | 8.3  | 687  | 16 Q8ABQ3 | Q8ABQ3 caulobacter  |
| 42 | 318.5 | 8.3  | 625  | 16 Q67320 | Q67320 aquifex aeo  |
| 43 | 315   | 8.2  | 789  | 16 Q8XX15 | Q8XX15 ralstonia s  |
| 44 | 314   | 8.2  | 673  | 16 Q8DDT0 | Q8DDT0 vibrio vuln  |
| 45 | 313.5 | 8.1  | 596  | 2 Q47423  | Q47423 escherichia  |

ALIGNMENTS

RESULT 1

Q9JVM4

ID Q9JVM4 PRELIMINARY; PRT; 761 AA.

AC Q9JVM4; 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Pilus secretin.

GN PILQ OR NMA0650.

OS Neisseria meningitidis (serogroup A).

OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;

OC Neisseriaceae; Neisseria.

OX NCBI\_TaxID=65699;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=22491 / Serogroup A / Serotype 4A;

RX MEDLINE=20222556; PubMed=10761919;

RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,

RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,

RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,

RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,

RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,

RA Whitehead S., Spratt B.G., Barrall B.G.;

RA "Complete DNA sequence of a serogroup A strain of Neisseria

meningitidis Z2491".

RL Nature 404:502-506(2000).

DR EMBL; ALI62753; CAB9338.1; -.

DR InterPro; IPR001775; Bac\_GSPD.

DR InterPro; IPR004846; GSPi/iiiprotein.

DR InterPro; IPR004845; GSPiiproteinC.

DR InterPro; IPR005644; NOLW-like.

DR Pfam; PF00263; GSPiI\_III.1.

DR Pfam; PF03958; GSPiI\_III.N; 1.

DR PRINTS; PR00811; BCTERIALGSPD.

DR PROSITE; PS00875; T2SP\_D; 1.

SQ Complete proteome.

Query Match 95.3%; Score 3666; DB 16; Length 761;

Best Local Similarity 96.2%; Pred. No. 3.4e-198;  
Matches 740; Conservative 2; Mismatches 19; Indels 8; Gaps 1;

QY 1 MNTKLTIIISGLFVATAAFAOTASAGNITDIKVSLLPNKQIKVSPDKIENVPTGVTSS 60  
DB 1 MNTKLTIIISGLFVATAAFAOTASAGNITDIKVSLLPNKQIKVSPDKIENVPTGVTSS 60

QY 61 PARIALDFEQTGISMDQOVLVYADPLSKISAQNSSRARLVNLNKPQGYNTEVGRNKV 120  
DB 61 PARIALDFEQTGISMDQOVLVYADPLSKISAQNSSRARLVNLNKPQGYNTEVGRNKV 120

QY 121 WIFINESDDTVSAPARPAVKAAPAAKQOQCRVTVQVRSIRIQTLYPGKTTAAAPFTES 180  
DB 121 WIFINESDDTVSAPARPAVKAAPAAKQOQCRVTVQVRSIRIQTLYPGKTTAAAPFTES 180

QY 181 VVSVAAPSPAKQOAAASAKOQTAAPAAKQOAAAPAKOTNIDPRKQKAGII 240  
DB 181 VVSVAAPSPAKQOAAASAKOQTAAPAAKQOAAAPAKOTNIDPRKQKAGII 232

QY 241 ELAALGFAGAGQPDISOQHDHIIIVTLKNHTLPTTLQRSLDVADFTPVQKVLKLNNDTOL 300  
DB 241 ELAALGFAGAGQPDISOQHDHIIIVTLKNHTLPTTLQRSLDVADFTPVQKVLKLNNDTOL 292

QY 301 IITAGNWLKNSAAGYFTFQVLPKKNLESQVGNNAKPTFTGRKISLDFOQVEIRTI 360  
DB 293 IITAGNWLKNSAAGYFTFQVLPKKNLESQVGNNAKPTFTGRKISLDFOQVEIRTI 352

QY 361 LQILAKESGMNIVASDSVNGKMTLSLKDVPMDQALDLVQARNLDMRQOQNI 420  
DB 363 LQILAKESGMNIVASDSVNGKMTLSLKDVPMDQALDLVQARNLDMRQOQNI 412

QY 421 LLAKDAFLQAEKDIADLGALYSNFKLYKNVEEFSIIRLNDADTTGNRTLVSGRS 480  
DB 413 LLAKDAFLQAEKDIADLGALYSNFKLYKNVEEFSIIRLNDADTTGNRTLVSGRS 472

QY 481 VLIDPATNTLIVTDRSVIEKFKRLIDELVPAQOQVMEIARIVEAADGFSRDLGVAFGAT 540  
DB 473 VLIDPATNTLIVTDRSVIEKFKRLIDELVPAQOQVMEIARIVEAADGFSRDLGVAFGAT 532

QY 541 GKKLKNDTSAFGWNVSFGGDDKWAETKINLPITAAANSISLVRASISGALNELSA 600  
DB 533 GKKLKNDTSAFGWNVSFGGDDKWAETKINLPITAAANSISLVRASISGALNELSA 592

QY 601 SESLKTTLANPRVLTONRKEAKIESGYEIPFTVTSIANGGSSTNTLKAVALGLVTVP 660  
DB 593 SESLKTTLANPRVLTONRKEAKIESGYEIPFTVTSIANGGSSTNTLKAVALGLVTVP 652

QY 661 NITPDGQIIMTVKINKDSPACASGNQTLICISTKNLNTQAMVNGGTLIVGGIYEEDNG 720  
DB 653 NITPDGQIIMTVKINKDSPACASGNQTLICISTKNLNTQAMVNGGTLIVGGIYEEDNG 712

QY 721 NTLTKVPLLDGIPVIGNLFKTRGKKTDRRELLIIFITPRINGTAGNSLRY 769  
DB 713 NTLTKVPLLDGIPVIGNLFKTRGKKTDRRELLIIFITPRINGTAGNSLRY 761

RESULT 2  
Q9ZHF3  
ID Q9ZHF3 PRELIMINARY; PRT; 765 AA.  
AC Q9ZHF3;  
DT 01-MAY-1999 (Tremblrel. 10, Created)  
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Secretin.  
GN Pilo.  
OS Neisseria meningitidis.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
OC Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=487;  
RN [1]  
RC STRAIN=H4/76;  
RX MEDLINE=98367129; PubMed=9701807;

Toujunt T., Caugant D.A., Dunham S.A., Koomey M.;  
Structure and function of repetitive sequence elements associated  
with a highly polymorphic domain of the *Neisseria meningitidis* Pilo  
protein.";  
Mol. Microbiol. 29:111-124(1998).  
EMBL; AF066056; AAC96097.1; -;  
InterPro; IPR001775; Bac\_GSPD.  
InterPro; IPR004846; GSP1/IIIProtein.  
InterPro; IPR005644; NoliW-like.  
Pfam; PF00263; GSP1\_III; 1.  
Pfam; PF03958; GSP1\_III; 1.  
PRINTS; PRO0811; BCTERIALGSPD.  
SQ SEQUENCE 766 AA; 82087 MW; A5190B3514404D6D CRC64;

Query Match 96.2%; Score 3094; DB 2; Length 766;  
Best Local Similarity 93.7%; Pred. No. 5.5e-166;  
Matches 641; Conservative 18; Mismatches 97; Indels 10; Gaps 3;

QY 1 MNTKLTIIISGLFVATAAFAOTASAGNITDIKVSLLPNKQIKVSPDKIENVPTGVTSS 60  
DB 1 MNTKLTIIISGLFVATAAFAOTASAGNITDIKVSLLPNKQIKVSPDKIENVPTGVTSS 60

QY 61 PARIALDFEQTGISMDQOVLVYADPLSKISAQNSSRARLVNLNKPQGYNTEVGRNKV 120  
DB 61 PARIALDFEQTGISMDQOVLVYADPLSKISAQNSSRARLVNLNKPQGYNTEVGRNKV 120

QY 121 WIFINESDDTVSAPARPAVKAAPAAKQOQCRVTVQVRSIRIQTLYPGKTTAAAPFTES 180  
DB 121 WIFINESDDTVSAPARPAVKAAPAAKQOQCRVTVQVRSIRIQTLYPGKTTAAAPFTES 180

QY 181 VVSVAAPSPAKQOAAASAKOQTAAPAAKQOAAAPAKOTNIDPRKQKAGII 240  
DB 181 VVSVAAPSPAKQOAAASAKOQTAAPAAKQOAAAPAKOTNIDPRKQKAGII 232

QY 233 DGNKAGIIELAALGFAGAGQPDISOQHDHIIIVTLKNHTLPTTLQRSLDVADFTPVQKVLK 292  
DB 241 DGNKAGIIELAALGFAGAGQPDISOQHDHIIIVTLKNHTLPTTLQRSLDVADFTPVQKVLK 300

QY 293 RLNDTQLIITAGNWLKNSAAGYFTFQVLPKKNLESQVGNNAKPTFTGRKISLD 352  
DB 301 RLNDTQLIITAGNWLKNSAAGYFTFQVLPKKNLESQVGNNAKPTFTGRKISLD 360

QY 353 QDVEIRTILOILAKESGMNIVASDSVNGKMTLSLKDVPMDQALDLVQARNLDMRQOQNI 412  
DB 361 QDVEIRTILOILAKESGMNIVASDSVNGKMTLSLKDVPMDQALDLVQARNLDMRQOQNI 420

QY 413 VNIAPRELLAKDAFLQAEKDIADLGALYSNFKLYKNVEEFSIIRLNDADTTGNRN 472  
DB 421 VNIAPRELLAKDAFLQAEKDIADLGALYSNFKLYKNVEEFSIIRLNDADTTGNRN 479

QY 473 TLVSGRSVLIDPATNTLIVTDRSVIEKFKRLIDELVPAQOQVMEIARIVEAADGFSRD 532  
DB 480 TLVSGRSVLIDPATNTLIVTDRSVIEKFKRLIDELVPAQOQVMEIARIVEAADGFSRD 539

QY 533 LGVAFGATGKKLKNDSAFGWNVSFGGDDKWAETKINLPITAAANSISLVRASISG 592  
DB 540 LGVAFGATGKKLKNDSAFGWNVSFGGDDKWAETKINLPITAAANSISLVRASISG 598

QY 593 ALNLELSASELSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSIANGGSSTNTLKA 652  
DB 599 ALNLELSASELSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSIANGGSSTNTLKA 658

QY 653 VLGLTVPTNITPDGQIIMTVKINKDSPACASGNQTLICISTKNLNTQAMVNGGTLIVG 712  
DB 659 VLGLTVPTNITPDGQIIMTVKINKDSPACASGNQTLICISTKNLNTQAMVNGGTLIVG 718

QY 713 GIYEDNGNTLTKVPLLDGIPVIGNLFKTRGKKTDRRELLIIFITPR 758  
DB 719 GIYEDNGNTLTKVPLLDGIPVIGNLFKTRGKKTDRRELLIIFITPR 764

RESULT 3  
Q50972



ID Q50972 PRELIMINARY; PRT; 720 AA.  
 AC Q50972;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE PiliQ.  
 GN PiliQ.  
 OS Neisseria gonorrhoeae.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=485;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MS11;  
 RX MEDLINE=96422484; PubMed=8825101;  
 RA Drake S.L., Koomsey M.;  
 RT "The product of the piliQ gene is essential for the biogenesis of type  
 IV pili in Neisseria gonorrhoeae";  
 RI Mol. Microbiol. 18:975-986(1995).  
 DR EMBL; U40396; AAC43603.1;  
 DR InterPro; IPR001775; Bac\_GSPD.  
 DR InterPro; IPR004846; GSPi/IIIproteins.  
 DR Pfam; PF00263; GSPiI\_III; 1.  
 DR Pfam; PF03958; GSPiI\_III; 1.  
 DR PRINTS; PR00811; BCTERIALGSPD.  
 DR SEQUENCE 720 AA; 77596 MW; A45BE2AD06DEB92B CRC64;  
 SQ  
 Query Match 77.7%; Score 2388; DB 2; Length 720;  
 Best Local Similarity 81.7%; Pred. No. 4.7e-160;  
 Matches 619; Conservative 21; Mismatches 78; Indels 40; Gaps 4;  
 QY 1 MNTKTKIISGLFVATAAQTASAGNITDIKVSLLPNKQIKVVSFDFKEIYNPTGFTVSS 60  
 DB 1 MNTKTKIISGLFVATAAQTASAGNITDIKVSLLPNKQIKVVSFDFKEIYNPTGFTVSS 60  
 QY 61 PARIALDFEQTGISMDQVLEAYDPLLSKISAAQNSRARLVNLNPKGQYNTVEVRGNKV 120  
 DB 61 PARIALDFEQTGISMDQVLEAYDPLLSKISAAQNSRARLVNLNPKGQYNTVEVRGNKV 120  
 QY 121 WIFINESDDTVSAPAPPAKAPAPAKQCGRTYVYRSIRIQLYPGKTTAAPTFS 180  
 DB 121 WIFINESDDTVSAPAPPAKAPAPAKQCGRTYVYRSIRIQLYPGKTTAAPTFS 180  
 QY 158 VVVSAPSPAKQQAASAKQQTAAAPAKQQAAPAKQQAAPAKQQAAPAKQQAAPAKQQA 240  
 DB 158 VVVSAPSPAKQQAASAKQQTAAAPAKQQAAPAKQQAAPAKQQAAPAKQQAAPAKQQA 240  
 QY 241 ELAALGFAGQPDISQGHDIIVTLKNHTLPTLQBSLDVADFKTPQKVTLLKRLNNDTOL 300  
 DB 241 ELAALGFAGQPDISQGHDIIVTLKNHTLPTLQBSLDVADFKTPQKVTLLKRLNNDTOL 300  
 QY 301 IITTAGNMELVNSAAGYFTFQVLPPKQNLSEGGVNNAPKFTTGRKISLDFQDVEIRTI 360  
 DB 301 IITTAGNMELVNSAAGYFTFQVLPPKQNLSEGGVNNAPKFTTGRKISLDFQDVEIRTI 360  
 QY 361 LOILAKESGMNIVASDSVNGKMTLSKQVDPDQALDLVQARNLDMROGGNIVNIAPRDE 420  
 DB 361 LOILAKESGMNIVASDSVNGKMTLSKQVDPDQALDLVQARNLDMROGGNIVNIAPRDE 420  
 QY 421 LLAKQKAFQAEKDIALGALYSQNFQKYNVEFRSILRLDNADTTGNRTLVSGRGS 480  
 DB 421 LLAKQKAFQAEKDIALGALYSQNFQKYNVEFRSILRLDNADTTGNRTLVSGRGS 480  
 QY 481 VLIDPATNTLIIVTDRSVIEKPRKILDELDPVPAQOVNIEARIVEAADGFSRDLGVKFGAT 540  
 DB 481 VLIDPATNTLIIVTDRSVIEKPRKILDELDPVPAQOVNIEARIVEAADGFSRDLGVKFGAT 540  
 QY 541 GRKKLKNBTSAFGWNSGFGGDDKWAETKINLPITAAANSISLVRAISSCALNLELSA 600  
 DB 541 GRKKLKNBTSAFGWNSGFGGDDKWAETKINLPITAAANSISLVRAISSCALNLELSA 600  
 QY 601 SESLSKTKTLANPRVLTONRKEAKIESGYEIPFTVTRSGGNSNTTEKKAVLGLVTVA 620  
 DB 601 SESLSKTKTLANPRVLTONRKEAKIESGYEIPFTVTRSGGNSNTTEKKAVLGLVTVA 620

DB 561 SESLSKTKTLANPRVLTONRKEAKIESGYEIPFTVTRSGGNSNTTEKKAVLGLVTVA 620  
 QY 661 NITPDGQIIMTVKINKDSPQACASGNTILCISTKNLNTQAMVNGGTLIVGGIYEEDNG 720  
 DB 621 NITPDGQIIMTVKINKDSPQACASGNTILCISTKNLNTQAMVNGGTLIVGGIYEENNG 680  
 QY 721 NLTUKVPLLDGPIVGNLFPKTRGKTKDRRELLIFITPR 758  
 DB 681 NLTUKVPLLDGPIVGNLFPKTRGKTKDRRELLIFITPR 718  
 RESULT 4  
 Q8XV60 PRELIMINARY; PRT; 714 AA.  
 AC Q8XV60;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Probable fimbrial type-4 assembly signal peptide protein.  
 GN PiliQ OR RSC2971 OR RS01326.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Ralstoniaceae; Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Arlat M., Billaut A., Brottier P., Camus J.C., Cattolico L.,  
 RA Chandler M., Choise N., Claudel-Renard C., Cunhac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502(2002).  
 DR EMBL; AL646073; CAD16680.1;  
 DR InterPro; IPR001775; Bac\_GSPD.  
 DR InterPro; IPR004846; GSPi/IIIproteins.  
 DR InterPro; IPR004845; GSPiIproteins.  
 DR InterPro; IPR005644; NOLW-like.  
 DR InterPro; IPR003522; SecYII-OMPG.  
 DR Pfam; PF00263; GSPiI\_III; 1.  
 DR Pfam; PF03958; GSPiI\_III; 1.  
 DR PRINTS; PR00811; BCTERIALGSPD.  
 DR PRINTS; PR01337; TYPE3OMGPROT.  
 DR PROSITE; PS00875; T2SP\_D; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 714 AA; 76611 MW; 94AE6FB5F8E5A070 CRC64;  
 Query Match 28.4%; Score 1091; DB 16; Length 714;  
 Best Local Similarity 34.8%; Pred. No. 3e-53;  
 Matches 270; Conservative 132; Mismatches 253; Indels 120; Gaps 18;  
 QY 8 IISGLFVATAFOTASAGNITDIKVSLLPNKQIKVVSF-DKEIVNPTGFTVSSPARIAL 66  
 DB 31 MVGGQALAQAI-TAPASNAVERVEQASTGESTVTVTLKDTPAQKPVFSTQOPARIAL 89  
 QY 67 DFQGTGISMDDQVLEAYDPLLSKISAAQNSRARLVNLNPKGQYNTVEVRGNKVIFINE 126  
 DB 90 DFQGTAPAQGRANYQVGGKLLRAASVVQIGDRTRVLDLARQSQYKSEIRGNQVLTGL- 148  
 QY 127 SDDTVSAPAPPAKAPAPAKQCGRTYVYRSIRIQLYPGKTTAAPTFSVSVSA 186  
 DB 149 -----AAPTASA----- 155  
 QY 187 PFSAPAKQAASAKQQTAAAPAKQQAAPAKQQAAPAKQQAAPAKQQAAPAKQQA 244  
 DB 156 -----TPVPT---FAAPATAGVERPSVRNIDFRGEELAGRVVDLST 196  
 QY 245 LGFAGQPDISQGHDIIVTLKNHTLPTLQBSLDVADFKTPQKVTLLKRLNNDTOLIIT 304  
 DB 245 LGFAGQPDISQGHDIIVTLKNHTLPTLQBSLDVADFKTPQKVTLLKRLNNDTOLIIT 304

Db 197 SNSA--INIAQQGNLVVDFAGATLPSLRFRFVDFGTPVQAMRATDNGTGARLVIEP 254  
 Qy 305 AGNMELVNKSAAPGYFTFQVLPPKQK---LESQGVNNAKPTFTGKISLDFQDVEIRTL 361  
 Db 255 RGNWQ--YSSYQTDQFVVEVRPTKEDPNKLISG-----PCYRGERSLNFQNDIRSL 307  
 Qy 362 QILAKESGMNIVASDYNKMTLSLKDPVWDQALDVMQAEKNDLMDROQGNVNIAPRDEL 421  
 Db 308 QVFADFTNLNITSDSVTGLTSLRSLKDPVWDQALDVMQAEKNDLMDROQGNVNIAPRDEL 367  
 Qy 422 LAKDAFLQAEKNDLADLQALYQNFQKLYKVEEFRSLRLDNADTTGN-----RNTLVSGR 478  
 Db 368 ATKEAELESQOQVTELEPLASQVFLNYQRAADVRLNL--LGTGGTAGGGAASRILSKR 426  
 Qy 479 GSVLIDPATNLIVTDTRSVIEKPKLDELDPVQAOVMIEARIVEADGSRDLGVKFG 538  
 Db 427 GSLTSDARTNLQFVSDIPSKLEEVQAFLLKIDIPVQVMIEARIVEADDTESRLGAKLG 486  
 Qy 539 ATGKKLNDTSFAFGW-----VNSGFGDDKQWGAETKINLP-----ITAAANSISLVR 587  
 Db 487 FPAK-----INGAGYGNNTYNSVSPVTKDATWDSAAVSLPANGINGVNAASVAVSLFN 540  
 Qy 588 AISSGALNLELSASELSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSIANGSSSTNT 647  
 Db 541 AGAGFLALELSALEADRGKTISSPRVVTADNIKALIEQGTVEPV---QOATSSGATSV 597  
 Qy 648 ELKKAFLGLTVTPNITPDQOIIMTVKINKDSPAOCASGNQTI--LCISTKNTLQAMVEN 705  
 Db 598 SFKANKLKEVTPKTIIPDGNVFLDVNKS-----LGTQITNGPAINTKHVQIQVLVEN 652  
 Qy 706 GTTLVGGIYEEDNGNTLTKVPLGDIPIVGNLFKTRGKKTDRRELLIFITPR 760  
 Db 653 GGTVVIGGIYQNTERTDVKVPLGDIPIVGNLFKSTAKTNDRTELLVFLTERVL 707

RESULT 5  
 Q9AM59 PRELIMINARY; PRT; 723 AA.  
 AC Q9AM59  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative outer membrane protein ComQ.  
 GN COMQ.  
 OS Acinetobacter sp. BD413.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Moraxellaceae; Acinetobacter.  
 OX NCBI\_TaxID=104611;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rosenblaser C., Averhoff B.  
 RT "Genes essential for the natural transformation process in  
 RL Acinetobacter sp. BD413.";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF329876; AAK0351.1; -  
 DR InterPro; IPR001775; Bac\_GSPD.  
 DR InterPro; IPR004846; GSP1/IIIProtein.  
 DR InterPro; IPR004845; GSP1/IIIProtein.  
 DR InterPro; IPR005644; NOLW-like.  
 DR Pfam; PF00263; GSP1\_III; 1.  
 DR Pfam; PF03958; GSP1\_III\_N; 1.  
 DR PRINTS; PR00811; BCTERIALGSPD.  
 DR PROSITE; PS00875; T2SP\_D; 1.  
 SQ SEQUENCE 723 AA; 78166 MW; E09AC8369907DAB9 CRC64;

Query Match 27.3%; Score 1050; DB 2; Length 723;  
 Best Local Similarity 32.1%; Pred. No. 6.2e-51;  
 Matches 251; Conservative 143; Mismatches 260; Indels 128; Gaps 17;

Qy 14 VATAAFQTASAG-NITDIKVSLLPNKQIKVKSFDKEIVNPTGFTVTSPPARIALDFEQTG 72  
 Db 27 VAVAIQVASAQVAITNVVPMNIPGQGTETRVNFMENGLPPQPAQVLEQPARILDFDKAG 86

Qy 73 ISWQOVLEVADPLLSKISAAQNSRRARLVNLNPKFGQVNTVRGNKVMIFINESDDTVS 132  
 Db 87 QKLAKSLIPVTNEASSIDVSSDDKARVVVNLKADGAPTTRVEGN---VFI----- 135  
 Qy 133 APARPAKAAAPAKAQOQCRTVYQVRSIRIOTLTPGKTTAAAPFTESVVSAPSPAK 192  
 Db 136 -----LKNINI----- 141  
 Qy 193 QQAAASAKOCTAAPAKQOATAAPAKOQAAAPAKOTNIDPRKDGKNAGIIEIAALGFAGQPD 252  
 Db 142 -----OSTATPV--ATSAFAPOQIG-----NIGFORGAQOGLVVIDLQGTNTPVD 186  
 Qy 253 ISQOHDHIIIVTLKNTLPTTLQSLRDLVAFKTPVQVKTLKRLNNDTQLIITTAGNHWELVN 312  
 Db 187 VQOQSKVVVFRPGIKIPHTLARRLNTTDFATPVASIDSYNDGNGVSIQSTGSEYEM- 245  
 Qy 313 KSAAPGYFTFQVLPPKQKLESGVGNNAKPTFTGKISLDFQDVEIRTLQILAKESGMNI 372  
 Db 246 AYQAEKNTLSLKRPOCNVRLSKNN--QNYTGKLSLDFQIEVRVQLLADFNFNINM 303  
 Qy 373 VASDSVNGKMTLSKDPVWDQALDVMQAEKNDLMDROQGNVNIAPRDELAKD-----KAF 428  
 Db 304 VTADSVQGNITLHLKDPVWDQALDIILKTKNLDKRRNGNVMIAPVSELIKSEDEAKAI 363  
 Qy 429 LQAEKDIADLQALYQNFQKLYKVEEFRSLRLDNADTTGNR-----T 473  
 Db 364 AQSTK-----LAPLOTEYILLYAKAADIIEKLTQGRNSGTSTNSNTNGSATVEPLGDSVGT 419  
 Qy 474 LVSGRSLVDPATNLIVTDTRSVIEKPKLDELDPVQAOVMIEARIVEADGSRDL 533  
 Db 420 LLSPRGTISLDPRTNLTINDTSQKIDQIRKMDLLDIPVKQVMIEARIVRATTSFKEM 479  
 Qy 534 GVNFQATGKKLNDTSAFG-----WGVNSGFGDDKQWGAET-----KINLPI---TAA 579  
 Db 480 GVKWGLISQGNINQNSNLLVGGSETTLNLRN-----PDSNGTYTIERPDNLNDLGVTTTG 535  
 Qy 580 ANSISL-VRAISSGALNLELSASELSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSI 638  
 Db 536 ASSIAFGLISLDFMLDLELSALQADGYGEVISTPKVMSADKQKAVATGVEVPYQSTTN 595  
 Qy 639 ANGSSNTNTEKXAVLGLTVTPNITPDQOIIMTVKINKDSPAOCASGNQTIICISTKNTLN 698  
 Db 596 SAAGTATTSFKEALLSLEVTNITPDGKILMELNANDINSYAQNGEAIL--NKNNTIN 653  
 Qy 699 TQAMVENGGLTVGGIYEEDNGNTLTKVPLGDIPIVGNLFKTRGKKTDRRELLIFITPR 758  
 Db 654 TNLVNNGETVVLGGVFEQTNNNAVTKVPLGDIPIVGNLFKTRGKKTDRRELLIFITPR 713  
 Qy 759 IM 760  
 Db 714 IV 715

RESULT 6  
 Q8EK21 PRELIMINARY; PRT; 684 AA.  
 AC Q8EK21  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Type IV plus biogenesis protein PilQ.  
 GN S00285.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadaceae; Shewanella.  
 OX NCBI\_TaxID=70863;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MR-1;  
 RX MEDLINE=22297686; PubMed=12368813;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
 RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,

## RESULT 7

Db 384 ARDLGAREGIARRQYDNTAVISGSSSTNNVSVINDGL--HNVPGLNPNLPGATTN 440  
 Qy 579 AANSLVRA:SSGALNELSASELSKTKLANPVLTONRKEAKIESGYEIPF-TVTS 637  
 Db 441 TAGSAYTLGRNFPALDELNLSAMQEGEVSVNPRIVTANQREGVIKQGREIGVWISG 500  
 Qy 638 IANGG--SSNTTELKAVJGLTVNPNITPDGQIIMTVKINKSPACAS-----GNOTILCI 692  
 Db 501 GGTGQSQANVQFKEVLELKLKTPITNDNRVFLNMNVKDEVARFINLPYLG--TVPEI 558  
 Qy 693 STPKNTQAWNGGTLVGGYEBDNGTTLTKVPLLDIPVIGNLPKTRGKTKDRRELL 752  
 Db 559 NRRENVAVLVDGETVTVIGGYEYFENDRESVAKVPLDGIPLGNLFKKRGRSKEAELL 618  
 Qy 753 IFITPRIMTAG 764  
 Db 619 VFVTPKVLRVAG 630

RESULT 8  
 Q8P5V7 PRELIMINARY; PRT; 648 AA.  
 AC Q8P5V7;  
 DT 01-OCT-2002 (Tremblrel. 22, Created)  
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Fimbrial assembly protein.  
 GN Xanthomonas campestris (pv. campestris).  
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 OX NCBI\_TaxID=340;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33913 / NCPPB 529;  
 RX MEDLINE=2202145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Camarotte G., Cannavari F., Cardozo J., Chambergo F., Chapina L.P.,  
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,  
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 RA Spinoza L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Setubal J.C., Kitajima J.P.;  
 RA "Comparison of the genomes of two Xanthomonas pathogens with differing  
 RT host specificities."  
 RL Nature 417:459-463(2002).  
 DR EMBL; AE012440; AAM42498.1; -;  
 DR InterPro; IPR001775; Bac GSPD.  
 DR InterPro; IPR004846; GSP1/IIIprotein.  
 DR InterPro; IPR004845; GSP1/IIprotein.  
 DR InterPro; IPR005644; NoW-like.  
 DR Pfam; PF00263; GSP1\_III; 1.  
 DR Pfam; PF03958; GSP1\_III; 1.  
 DR PRINTS; PR00811; BCTERIALGSPD.  
 DR PROSITE; PS00875; TSP\_D; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 648 AA; 68993 MW; 82C674FCB332A1CB CRC64;

Query Match 23.6%; Score 910; DB 16; Length 648;  
 Best Local Similarity 34.0%; Pred. No. 4e-43;  
 Matches 216; Conservative 120; Mismatches 215; Indels 84; Gaps 12;  
 195 AAGSAKQQTAAAPAKQTAAPAKQAAPAKQTNIDFRKDGNGAGIIEALGAFAGQPDIS 254

Db 28 ALASGSSFAAALQAQAPAK-TAPASLAVSKIPIKRGDDGAGRLILOFDOGASPDUR 86  
 Qy 255 QQHDHIIIVLKNHTLTPTTLORSIDVADFKTPQKVTLKLNNDTQLIITAGNWLNVKS 314  
 Db 87 TQGDNVLDVSNARLPAELQRPUNVTFATPVQREVFKSGGSQLVSTKGAFDSLAVQ 146  
 Qy 315 AAGGYFTFQVLPKQNLSEGGVN-----NAPKPTFGKISLDPODVEIRTLQIL 364  
 Db 147 TGNEY-VVEITPRKGPVAGVSVSAVTAQAAGYSGREVTFNFQDPVPTVLOLI 205  
 Qy 365 AKSGGNIVASDSVNGKMTLSLKVDPNDQALDVMQARNLDMQOQGNVNIAPRDEL--L 422  
 Db 206 ABEENLIVASDVTQGVNVTURLMNVDPWDQALDVLRAKGLDKRDDGVVWVAQPELAKF 265  
 Qy 423 AKDK-----AFLOAEKDIADLGALYSONFOLKYKNVEEPRSLRLDNADTTGNR----- 471  
 Db 266 EQDKEDARIAIENREDLITDYQINYNHAAVIFKALTEAKGIGGGGGGGGGGAGGQ 325  
 Qy 472 -NTLVSGRGSVLIDPATNTLIVTDRSVIEKFKLIDELDVPAQOQMIARIVEAADGFS 530  
 Db 326 DNGFLSPRGRLVADERTNTLMISDIPKKVQAMRELISHIDRPVDQVLIESRIVADTFPA 385  
 Qy 531 RDLGVKFGATGKKLKNDTSAFGWVNSGFGGDDKMGAEKINLPITAAANSLSVRA-- 588  
 Db 386 RDLGARFGVTG-----ATGGILSG-----SLESNNVNLNTSAQSRLEQANGGQ 429  
 Qy 589 -----ISSGA-----INLELSASELSKTKTLANP 613  
 Db 430 VTTLPALHFPGLNVDLGAGGFTNSGAAGLAYTLGSHFNLDIELSAWQEGREVVSNP 489  
 Qy 614 RVLQNRKEAKIESGYEIPFTVTS---IANGSGSTNELKAVLGLTVTPNITPDGQIM 670  
 Db 490 RIVTANQREGVIKQREIGYVTVISGAVAGGSGQANVQFKEVLELKLKAVTPTIINDNRVFL 549  
 Qy 671 TVKINKSPACASGQ--TILCISTKNLNTQAVENGGLIVGGIYEENGNTLTQVPL 728  
 Db 550 NMNVKDEVARFITLPQVGTPEINREVNATVLVADGETVWIGGYEFTDRESVAKVPF 609  
 Qy 729 LGDIPVIGNLPKTRGKTKDRRELLIFITPRIMTAG 763  
 Db 610 LGDIPVIGNLPKTRGKTKDRRELLIFITPRIMTAG 763

RESULT 9  
 Q8P5V7 PRELIMINARY; PRT; 637 AA.  
 AC Q8P5V7;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Fimbrial assembly protein.  
 GN XF0373.  
 OS Xylolla fastidiosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xylolla.  
 OX NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9a5C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carret H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,  
 RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hohnsbeil J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kurama E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.R., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,

RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Teai S.M., Tshako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Seubai J.C.,  
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"  
 RL Nature 406:151-159(2000).  
 DR EMBL; AE003888; AAF83183.1; -  
 DR InterPro; IPR001775; Bac GSPD.  
 DR InterPro; IPR004846; GSP11/IIIProtein.  
 DR InterPro; IPR004845; GSP11/IIIProtein.  
 DR InterPro; IPR005644; NOLW-like.  
 DR Pfam; PF00263; GSP11\_III; 1.  
 DR Pfam; PF03958; GSP11\_III; 1.  
 DR PRINTS; PR00811; BCTERIALGSPD.  
 DR PROSITE; PS00875; T2SP\_D; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 637 AA; 68901 MW; 8DBFCD7EE464EC9 CRC64;  
  
 Query Match 33.6%; Score 907; DB 16; Length 637;  
 Best Local Similarity 34.5%; Pred. No. 5.8e-43;  
 Matches 203; Conservative 127; Mismatches 221; Indels 38; Gaps 11;  
  
 QY 207 AKQTAAPAKQQAAPAKQTNIDPRKDGKNAGIIEAALGFAGQPDISOQHDHIIIVTLKN 266  
 DB 48 AENKQATPVKPVANAPUSVKIDFKRGDDSGRLILKFDQGGATPDLRTGGTVLVLGT 107  
  
 QY 267 HTPTTLQRLSDVADFTPVQKVTLLKRLNDTQIIITTAGNWLNVKSAAPGYFTFQVLP 326  
 DB 108 AVLPTVLQRIQNVVDFATPQRIQDAKPMGKAQLVLTGKGFSLAYQTGDEY-VVEIIVP 166  
  
 QY 327 KQKNLESQGV-----NNAPKT---FTGRKISLDFQDVEIRTLQILAKESGMNIVAS 375  
 DB 167 KGEAANGGAIPTESVSKSAKSAQYSGRPVTFNFDQVPTVTLQIADENINIVAS 226  
  
 QY 376 DSVNGKWTLSKDVDPWQDALDLVQARNLDMROQGNIVNIAPRDELLAKDFAFLQAEKDI 435  
 DB 227 DTVOGNVTLALINVPWQDALDLVRAKGLDKRDEKVIWVAPQOELAKYQEKEDARIAI 286  
  
 QY 436 ADLGALYSQNFQKLYKNVEE-FRSLRLDNADTTGN-----RNTLVSGRGSVLDPAT 487  
 DB 287 ENREGLITDVQINHYSAIVFKAITEAKMGWGGNGTNNNDAAFLSPRGVLADERT 346  
  
 QY 488 NTLIVTDRSVIEKFRKLIDELDPVQAQVMIEARIVEAADGFRDLGVKFGATGKXKLN 547  
 DB 347 NTLMISDIPKVAQMRTLIQHIDRPVQVLEIGRIVIATDSFARDLGAKFGVGAASRPSD 406  
  
 QY 548 DTSAFGWNVSGFGGDDKWAETKINLPITAAANSISLVPAISSGA-----LNLEL 598  
 DB 407 NTATIGSHVTT---ADSSATGATGLNVDLGGPTNTASV--LPSLAYTLGPKFNLDLEL 461  
  
 QY 599 SASELSKTKTLANPRVLTQNRKEAKIESGYEIPP-TVT-SIANGSGSTNTELKAVLGL 656  
 DB 462 SALQOESRGEVSNPRIVTANQREGYIKQCKEIGYVITGGVAGGQATPNVQKEAVLEL 521  
  
 QY 657 TVTPNITPDQQLIMTVINKSPAQ--CASGNQTLICISNTKNLQTMQWENGTLVGGI 714  
 DB 522 RVTPFTINDNRVFNMTVKKDEIDQMIIANFGFTVPLNKRREINTAVLVDDGQTVVIGV 581  
  
 QY 715 YEEDNGNTAKVPLGLGIPVIGNLFTKRGKTKDRRELLIETPRIMGTA 763  
 DB 582 YEFSDRNSVSKVPLGDPVPLGNLFKRGDKQKQAEILLIIVTPKLEVA 630

RESULT 10

Q9KXV0  
 ID Q9KXV0 PRELIMINARY; PRT; 578 AA.  
 AC Q9KXV0;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Fimbrial assembly protein.  
 GN VC2630.  
 OS *Vibrio cholerae*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; *Vibrio*.  
 OX NCBI\_TaxID=666;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=El Tor N16961 / Serotype O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Colwell R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*,"  
 RL Nature 406:477-493(2000).  
 DR EMBL; AE004329; AAF95771.1; -  
 DR TIGR; VC2630; -  
 DR InterPro; IPR001775; Bac GSPD.  
 DR InterPro; IPR004846; GSP11/IIIProtein.  
 DR InterPro; IPR004845; GSP11/IIIProtein.  
 DR InterPro; IPR005644; NOLW-like.  
 DR InterPro; IPR003522; SecIII\_OMP.  
 DR Pfam; PF00263; GSP11\_III; 1.  
 DR Pfam; PF03958; GSP11\_III; 1.  
 DR PRINTS; PR00811; BCTERIALGSPD.  
 DR PRINTS; PR01337; TYPE3ONGPROT.  
 DR PROSITE; PS00875; T2SP\_D; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 578 AA; 63248 MW; DB12A5E8DCA7F44C CRC64;  
  
 Query Match 21.0%; Score 809.5; DB 16; Length 578;  
 Best Local Similarity 34.3%; Pred. No. 1.6e-37;  
 Matches 191; Conservative 128; Mismatches 205; Indels 33; Gaps 14;  
  
 QY 218 QAAAPAKQTNIDPRKDGKNAGIIEAALGFAGQPDISOQHDHIIIVTLKNHTLPTTLQRLS 277  
 DB 39 ESATANQLENIDFRVKNKEAAVLIVELASPSAVDVQKVEGLSIELLKTVDADKLVL 98  
  
 QY 278 DVADFTPVQKVTLLKRLNDTQIIITTAGNWLNVKSAAPGYFTFQVLPKKNLESQGVN 337  
 DB 99 DYKDFSTPVESVEVPRKEPSTQVTVVDG--EFQHDYTLKGYLEWISKLKADE---K 152  
  
 QY 338 NAKPTF---TGRKISLDFQDVEIRTLQILAKESGMNIVASDSVNGKMTLSIKQVWPQQA 394  
 DB 153 PKPKSVLKEGKLIISINFQDIPVRNVLQIADYNGFNLVSDSVVGNUTLRLDGVWQOV 212  
  
 QY 395 LDLVNQARNLDMROQGNIVNIAPRDELLAKDFAFLQAEKDIADLGALYSQNFQKLYKNVE 454  
 DB 213 LDIIQLVQKGLDKRVDGNVILIAPKEELDRLREKALEKARLAEGLDKLSEIKINFAKAS 272  
  
 QY 455 EFRSLRLDNADTTQNRNTLVSGRGSVLIDPATNTLIIVTDRSVIEKFRKLIDELDPQA 514  
 DB 273 DTAAMI-----GGEQGNV-MLSERGISIDERTNSLLIRELPDNTAVIREIESLIDIVK 326  
  
 QY 515 QVMIEARIVEAADGFRDLGVKFGATGKXKLNKNTDSAFQWGVNS-----GFGDDDKWGA 569  
 DB 327 QVQIEARIVTVKEGNLEELGVWGVNS---TNGSHSVGGSTESNLWKGLLADDEFPVD 382  
  
 QY 570 --TKINLPITAA-ANSISL-VRAISSGA-LNLELSASELSKTKTLANPRVLTQNRKEAK 624  
 DB 383 EFLNVNLASTANASSIAFQVAKLGSGLDLDELALQNESKAEIISPRITTNKQPAY 442



Db 474 QARYGRKRVFEFQDIDIONLLRVIAISKKNIVADVSGKVTIRLNVPWQALDVL 533  
 Qy 400 QARNLDVRQGNIVNIAP-----RDELLAKDAFLQAEDIAADGALYSQNFQKLYKNV 453  
 Db 534 RTKALGKEEFGNIIRIAPLKTLEEARLRQERKKSQQQEDL--MYNLLPVYAVA---- 587  
 Qy 454 EERSILRLDONADTTGNRLTVSGRGSVLDPATNTLIVTDRSVIEKFKRLIDELDVPA 513  
 Db 588 -----ADWAARVKVDLSEGSVTVQDQTNVLIVKDRSNTERARSLVRSLDTQT 636  
 Qy 514 QQVMIEARIEAAGDFSRDLGVKFG-----ATGKKK-----LKNDDSAFGWGVNSGFGGD 563  
 Db 637 PQVLESRIEAVNTSFRSLSGVQWGGARAGQATGNSGLIFPNLAVTGGVGTGAGLP 696  
 Qy 564 DKWGAETKINLPITAAANSLVRAISSG-----LNLELSASELSKTKTLANPRVLT 617  
 Db 697 D--NPNFAVNLPTGTGGVGGANGFTFGAGGALQNLRLSAAENEGSVKTSAPKVT 753  
 Qy 618 QNRKEAKIESGIEPFTVTISANGSGSTNTELKAVLGLTVTPNITPDGOIIMTVKINKD 677  
 Db 754 LDNNTARINQGVSPFQTS-----AQCWTFVEARLSLEVTTHITQDGSVLMSINANN 809  
 Qy 678 SPAQASGNQTLICISTKNLNTQAMVNGGTLVGGIYEDNGNTLTKVPLLDGIPVGN 737  
 Db 810 QPQFSSTGANGQPSIQKEANTQVLKXGDDTVIGGIYVRRGATQVNSVFLSRIPVLGL 869  
 Qy 738 LFKTRGKKTDRRELLIIFITPRIM 760  
 Db 870 LFKNSETDRQELLIFITPRIL 892

RESULT 13

Q82206 Q82206 PRELIMINARY; PRT; 412 AA.  
 AC Q82206;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Type II secretion system protein.  
 GN HOFO OR STY4308.  
 OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Feitwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrall B.G.;  
 RA "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar Typhi CT18";  
 RL Nature 413:848-852(2001).  
 DR EMBL; AL627281; CAD08126.1;  
 DR InterPro; IPR001775; Bac\_GSPD.  
 DR InterPro; IPR004846; GSPII/IIIprotein.  
 DR InterPro; IPR004845; GSPIIproteinC.  
 DR InterPro; IPR005644; NOLW-like.  
 DR InterPro; IPR003522; SecIII-OMPG.  
 DR Pfam; PF00263; GSPII\_III; 1.  
 DR Pfam; PF03958; GSPII\_III; 1.  
 DR PRINTS; PR00811; BCTERIALGSPD.  
 DR PROSITE; PRO1337; TYPE3OMGPROT.  
 DR PROSITE; PS00875; T2SP\_D; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 412 AA; 44832 MW; B16155BC811A47DD CRC64;

Query Match 14.9%; Score 574.5; DB 16; Length 412;  
 Best Local Similarity 32.8%; Pred. No. 1.6e-24;  
 Matches 139; Conservative 94; Mismatches 148; Indels 43; Gaps 10;

Qy 347 KISLDFQVQVETILQILAKESGMNIVASDSVNGKWTLSLKDVPWQALDVLVQARNLDM 406  
 Db 23 KVTLVVDDVPVQVLOALAEQERQNLVSPDVSGLTSLHLTDVPWKQALQTVVNSAGLVL 82  
 Qy 407 ROQGNIVNIA-----PRDELLAKDAFLQAEDIAADGALYSQNFQKLYKNVEEFSILRL 462  
 Db 83 RQEGNILVHVSQAWQKEHSARQDAERLRQANL-----PLENRSISLOVADAGELAK---- 134  
 Qy 463 DNADTTGNRLTVSGRGSVLDPATNTLIVTDRSVIEKFKRLIDELDVPAQVMIEARI 522  
 Db 135 -----AGEK--LLSAKGTIMVDKRTNRLLLRNRALAELEKWKWSQMDLPVAQVELAAHI 187  
 Qy 523 VEADAGDFSRDLGVKFGATGKKLKNDDTSAFGWGVNSGFGGDKWGAETKINLPITAAANS 582  
 Db 188 VTINEKSLRELGVK-----TLADATQAGAVGDVTTLSSD-----LSVAAATSR 231  
 Qy 583 ISL-VRAISSGALNLELSASELSKTKTLANPRVLTQNRKEAKIESGIEIPTVTSIANG 641  
 Db 232 VGFNIGRINGRLDLLELSALEQKQQLDIASPRLLASHLQFASIKQSGEIPYQVSSGESG 291  
 Qy 642 GSGTNTLTKKAVLGLTVTPNITPDGOIIMTVKINKDSPAQC---ASGNQTLICISTKNLN 698  
 Db 292 --ATSVEKFAVLGMEVPTVLQGRIRLKLHISQNVFGVLQADGE--VLAIKQIE 347  
 Qy 699 TQAMVNGGTLVGGIYEDNGNTLTKVPLLDGIPVGNLTKRKTDRRELLIIFITPR 758  
 Db 348 TQVEVKSETLALGGIFSRKKNKSGSDSVPLLDGIDIPWLQGLFRHDKGDEKRELVVFIPTPR 407  
 Qy 759 IMGT 762  
 Db 408 LVAT 411

RESULT 14

Q82LK2 Q82LK2 PRELIMINARY; PRT; 412 AA.  
 AC Q82LK2;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative transport protein, possibly in biosynthesis of type IV  
 DE pilin.  
 DE HOFO OR STM3488.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Ryan E., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Waterston R., Wilson R.K.;  
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 RT LT2";  
 RL Nature 413:852-856(2001).  
 DR EMBL; AE008860; AAL22350.1;  
 DR InterPro; IPR001775; Bac\_GSPD.  
 DR InterPro; IPR004846; GSPII/IIIprotein.  
 DR InterPro; IPR004845; GSPIIproteinC.  
 DR InterPro; IPR005644; NOLW-like.  
 DR InterPro; IPR003522; SecIII-OMPG.  
 DR Pfam; PF00263; GSPII\_III; 1.  
 DR Pfam; PF03958; GSPII\_III; 1.  
 DR PRINTS; PR00811; BCTERIALGSPD.  
 DR PROSITE; PS00875; T2SP\_D; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 412 AA; 44832 MW; B16155BC811A47DD CRC64;



DR PRINTS; PRO1337; TYPE3ONGPROT.  
DR PROSITE; PS00875; T2SP D; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 412 AA; 44869 MW; AA5788AB47A20C5B CRC64;

Query Match 14.8%; Score 571.5; DB 16; Length 412;  
Best Local Similarity 33.3%; Pred. No. 2.4e-24; Indels 35; Gaps 10;  
Matches 140; Conservative 91; Mismatches 154;

QY 347 KISLPQDVEIRITLIQILAKESGMNIVASDSVNGKMTLSKDVDPDQALDVMQARNLDM 406  
DB 23 KVTLVVDDVPVQVQLTAEQERQNLVSPDVSGLTSLHLTDVPMKQALQTVNSAGLVL 82

QY 407 RQGNINWAPDELLAKDKFLQAEKDIADIGALYSQNFQKYKNVEFRSILRLDNAD 466  
DB 83 RQGNILVH--HSQAWQKEHS---AQDAERL--RFOANLPLENRSI-----SLQVADAGE 131

QY 467 TTGNRNNTLVSGRGSVLIDPATNTLIIVTDRSVIEKFRKLIDELDPVPAQQMIEARIVPAA 526  
DB 132 LAKAGEKLLSAGTIVMDKRTNRLLRDNRAALAELEKVVWSQMDLPVAGVELAAHIVTIN 191

QY 527 DQFSRLGVKFGATGKKLKNDSAFSGWVNSGFGGDDKMGAEKTKINLPITAAANSISL- 585  
DB 192 EKSLRELGVK-----TLADATQAGSGVGVTLTSGD-----LSVAAATSRVGFN 235

QY 586 VRAISSGALNLELSASELSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSIANGSGST 645  
DB 236 IGRINGRLDLELSALEQKQLDIIASPRLLASHLOPASIKQSGEIPYQVSSGESG--AT 293

QY 646 NTELKKAIVGLTVTPNITPDGQIIMTVKINKSPACQ---ASGNQITILCISTKNTQAM 702  
DB 294 SVEFKEAVLGMEVTPTVLQKGRIRLKLHISVNGVPGVQVLOQADGE--VLADIKOEIE 351

QY 703 VENGGLTIVGGIYEDNGNTLTKVPLLDGIPVIGNLFKTRGKKTDRRELLIFITPRIMGT 762  
DB 352 VKSGEILALGGISFRKNKSGSDSVPLLDGIPWLGQIFRHDGKEDERRELWVITPRVAT 411

RESULT 15  
Q8CVW9 ID Q8CVW9 PRELIMINARY; PRT; 412 AA.  
AC Q8CVW9;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DE Protein transport protein hofQ precursor.  
GN HOFQ OR C4161.  
OS Escherichia coli O6.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=217992;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;  
RX MEDLINE=22388234; PubMed=12471157;  
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
RT "Extensive mosaic structure revealed by the complete genome sequence  
of uropathogenic Escherichia coli.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
DR EMBL; AE016768; AA82599.1; -;  
KW Complete proteome.  
SQ SEQUENCE 412 AA; 44781 MW; 78416A4B76D6888C CRC64;

Query Match 14.8%; Score 570.5; DB 16; Length 412;  
Best Local Similarity 33.3%; Pred. No. 2.7e-24; Indels 41; Gaps 11;  
Matches 142; Conservative 94; Mismatches 147;

QY 346 RKISLDFQDVEIRITLIQILAKESGMNIVASDSVNGKMTLSKDVDPDQALDVMQARNL 405  
DB 22 QKVTLMVDDVPVQVQLTAEQERQNLVSPDVSGLTSLHLTDVPMKQALQTVVKSAGLI 81

QY 406 MRQGNIV---NIAPRELLAKDKFLQAEKDIADIGALYSQNFQKYKNVEFRSILRL 462  
DB 82 TROEGNILSVHSAWQNDNIARQEA--EQARAQNL--PLENRNITLQYADAGELAK---- 134

QY 463 DNADTTGNRNNTLVSGRGSVLIDPATNTLIIVTDRSVIEKFRKLIDELDPVPAQQMIEARI 522  
DB 135 -----AGEK--LLSAKSGMTVDKTKNRLLRDNKLTALSTLEQVWSQMDLPVQGVLSAHI 187

QY 523 VEADGFSRLGVKFGATGKKLKNDSAFSGWVNSGFGGDDKMGAEKTKINLPITAAANS 582  
DB 188 VTINEKSLRELGVK-----TLADAAQAGVGQVTLTSGD-----LSVATATTH 231

QY 583 ISL-VRAISSGALNLELSASELSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSIANG 641  
DB 232 IGRINGRLDLELSALEQKQLDIIASPRLLASHLOPASIKQSGEIPYQVSSGESG 291

QY 642 GSSTNTLTKKAIVGLTVTPNITPDGQIIMTVKINKSPACQ---ASGNQITILCISTKNTLN 698  
DB 292 --ATSVEFKEAVLGMEVTPTVLQKGRIRLKLHISVNGVPGVQVLOQADGE--VLADIKOEIE 347

QY 699 TOAMVENGGLTIVGGIYEDNGNTLTKVPLLDGIPVIGNLFKTRGKKTDRRELLIFITPR 758  
DB 348 TOVEVKSGEILALGGISFRKNKSGSDSVPLLDGIPWLGQIFRHDGKEDERRELWVITPR 407

QY 759 IMGT 762  
DB 408 LVSS 411

RESULT 16  
Q8X818 ID Q8X818 PRELIMINARY; PRT; 412 AA.  
AC Q8X818;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Putative transport protein.  
GN HOFQ OR Z4744 OR ECS4233.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11208551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grotbeck E.J., Davis N.W., Lim A., Dimallanta E.T., Potamousis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
RL Nature 409:529-533 (2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Onosubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Lida T., Takami H., Honda T., Sasakiwa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:11-22 (2001).  
DR EMBL; AE005561; AAG58491.1; -;  
DR EMBL; AF002565; BAB37656.1; -;  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSPF/IIIProtein.  
DR InterPro; IPR004845; GSPF/IIIProtein.  
DR InterPro; IPR005644; NoIw-like.  
DR InterPro; IPR003522; SecIII\_OMP.

DR Pfam; PF00263; GSP11 III; 1.  
DR Pfam; PF03958; GSP11-III; 1.  
DR PRINTS; PRO0811; BCTERIALGSPD.  
DR PRINTS; PRO1337; TYPE3OMGPROT.  
DR PROSITE; PS00875; T2SP\_D; 1.  
KW Complete proteome.  
SQ SEQUENCE 412 AA; 44744 MW; C34B8C740A174D3E CRC64;  
  
Query Match 14.78; Score 567.5; DB 16; Length 412;  
Best Local Similarity 33.51; Pred. No. 4e-24;  
Matches 142; Conservative 94; Mismatches 147; Indels 41; Gaps 11;  
  
QY 346 RKISLDQDVEIRTIQLAKESGMNIVASDSVNGKMTLSKDPVMDQALDVMQARNLDMRQOQGNVNIAPRDE 405  
DB 22 QKVTLMVDDVPVAVLQALAEQEKLNIVSPDVSGTVSLHLTDVPMKQALQTVVKSAGLI 81  
  
QY 406 MRQOQGNV---NTAPRELLAKDKAFLOAKDADIGALYSQNFQKYKVEFRSILRL 462  
DB 82 TROEGNLSVHSIAWQNDNIAOEAE--EQARAQANL--PLENRNITLQYADAGLAK---- 134  
  
QY 463 DNADTTGNRNTLVSGRGSVLIDPATNTLIVTDRSVIEKFRKLIDELDPVPAQQWMEARI 522  
DB 135 -----AGEK--LLSAKSGMTVDKRTNLLRDNKNTALSALAEQWQMDLPVGQVELSAHI 187  
  
QY 523 VERADGSRDLGVKFGATGKKLKNIDTSAFSGVNGVSGFGDDKGAETKINLPITAAANS 592  
DB 188 VTINEKSLRELGVKM-----TLADAQHAGGVGVQVITLGS-----LSVATATTH 231  
  
QY 583 ISL-VRAISSGALNLELSASESLSKTTLANPRVLTQNRKEAKIESGYEPFTVTIANG 641  
DB 232 VGNIGSINGRLDLLELSALEQKQOLDI IASPLLASHLOPASIKQSEIPIQVSSGSG 291  
  
QY 642 GSSTNTLKVAVGLTVPNTITPDGQIIMTVKINKSPAC--ASGNQITLCISTKNLN 698  
DB 292 --ATSVFKEAVLGMETPTVLQGRIRLKLHISQNVPGVLOQADGE--VLAIDKQEI 347  
  
QY 699 TOAMVENGTLVIGVIEEDNGNTLVKPLGIDIPVGNLFTKRGKKTDRRELLIFITPR 758  
DB 348 TQVEVKSGETALGCIETRNKSKQSDSVPLLGIDIPWFGQLFRHDGKEDERRELVVITPR 407  
  
QY 759 IMGT 762  
DB 408 LVSS 411  
  
RESULT 17  
Q8ZJF8 PRELIMINARY; PRT; 374 AA.  
AC Q8ZJF8  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Putative membrane transport protein (Putative transport portein).  
GN YF00150 OR HOFG OR Y3932.  
OS Yersinia pestis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersinia.  
OX NCBI\_TaxID=632;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CO-92 / Biovar Orientalis;  
RX MEDLINE=21470413; PubMed=11586360;  
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,  
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdono-Tarraga A.M.,  
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,  
RA Leather S., Moulé S., Oyston P.C.F., Quail M., Rutherford K.,  
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;  
RT "Genome sequence of Yersinia pestis, the causative agent of plague."  
RN Nature 413:523-527(2001).  
RN [2]  
RP SEQUENCE FROM N.A.

RC STRAIN=KIMS / Biovar Mediaevalis;  
RX MEDLINE=22137863; PubMed=12142430;  
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,  
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,  
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,  
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,  
RA Perty R.D.;  
RT "Genome sequence of Yersinia pestis KIM."  
RL J. Bacteriol. 184:4601-4611(2002).  
DR EMBL; AJ414141; CAC89013.1; -;  
DR EMBL; AE013597; AAM87476.1; -;  
DR InterPro; IPR001775; Bac GSPD.  
DR InterPro; IPR004846; GSP11/III protein.  
DR InterPro; IPR005644; NOLW-like.  
DR InterPro; IPR003522; SecIII\_OMPG.  
DR Pfam; PF00263; GSP11\_III; 1.  
DR Pfam; PF03958; GSP11-III; 1.  
DR PRINTS; PRO0811; BCTERIALGSPD.  
DR PRINTS; PRO1337; TYPE3OMGPROT.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 374 AA; 41171 MW; 59945A56121328FA CRC64;  
  
Query Match 14.11; Score 543; DB 16; Length 374;  
Best Local Similarity 31.6%; Pred. No. 8.3e-23;  
Matches 127; Conservative 91; Mismatches 152; Indels 32; Gaps 7;  
  
QY 361 LQILAKESGMNIVASDSVNGKMTLSKDPVMDQALDVMQARNLDMRQOQGNVNIAPRDE 420  
DB 1 MQALADYRQLNLTITTTGGNLSRLIEVPWEQALAILRMGRKKAEREGTVMVFTQE 60  
  
QY 421 LIAKDKAFLOAKDADIGALYSQNFQKYKVEFRSILRLDNADTTGNRNTLVSGRGS 480  
DB 61 IOERQ-----QRTKQAAPEALANLTALQYANAEQV-----ADSLDPL--QGGLSLPS 109  
  
QY 481 VLIDPATNTLIVTDRSVIEKFRKLIDELDPVPAQQWMEARIVEAADGFSRDLGVKFGAT 540  
DB 110 VVADKRTNTLIRDTPASLALKWLIEMDLPQQVQISAHITVITSSDQLQELGYR---- 165  
  
QY 541 GKXKLNKDTSAFGVNGVSGFGDDKGAETKINLPITAAANSISL-VRAISSGALNLELS 599  
DB 166 -----WGMGEGKGNTRALRINDFNVLPLPNSAASVGFHVARIGRLLLELS 212  
  
QY 600 ASELSKTKTLANPRVLTQNRKEAKIESGYEPFTVTSIANGSGSSTNTLKVAVGLT 659  
DB 213 ALEQENQVDITIASPLITSHOOTASIKOGSDIPIYVS--RGKKEAAAEFEKVEALUGMEV 270  
  
QY 660 PNITPDGQIIMTVKINKSPA--QCASGNQITLCISTKNLNTQAMVENGTLVIGVIEED 718  
DB 271 PKILRNGKIILDKISQNMPIGIIKRGESMLLDKQIKQIKQTVNDGETIVLGGIFQOK 330  
  
QY 719 NGNTITKVPILGQDIPVGNLFTKRGKKTDRRELLIFITPRIM 760  
DB 331 KRQSVNKVPLLIADIPILGAMFRQDTQQOSRRELVIITPKLI 372  
  
RESULT 18  
Q9CLK3 PRELIMINARY; PRT; 444 AA.  
AC Q9CLK3  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Come.  
GN COME OR PM1225.  
OS Pasteurella multocida.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Pasteurella.  
OX NCBI\_TaxID=747;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Pm70;  
RX MEDLINE=21145866; PubMed=11248100;  
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;

RT "Complete genomic sequence of Pasteurella multocida pm70.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).  
 DR EMBL: AE006162; AA03309.1; -.  
 DR InterPro: IPR001775; Bac GSPD.  
 DR InterPro: IPR004846; GSPII/IIprotein.  
 DR InterPro: IPR004845; GSPIIproteinC.  
 DR InterPro: IPR005644; NOLW-like.  
 DR Pfam: PF00263; GSPII\_III; 1.  
 DR Pfam: PF03958; GSPII\_III; 1.  
 DR PRINTS: PR00811; BCTERIALGSPD.  
 DR PROSITE: PS00875; T2SP\_D; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 444 AA; 49206 MW; 6683BA089862633 CRC64;

Query Match 12.3%; Score 473.5; DB 16; Length 444;  
 Best Local Similarity 29.0%; Pred. No. 8.7e-19;  
 Matches 130; Conservative 99; Mismatches 155; Indels 65; Gaps 14;

QY 346 RKISLDF-----QDVE-----IRTIQLAKESGMNIVASDVNGKM 382  
 DB 6 RKISLVFLGCVAYGSSQDAEHFVLRKQAPLVEMQLYLQHQHLLIDHLEGTL 65  
 QY 383 TLSLKDVPDQALDVMQARNLMDROGNIWNIAPRDELLAKKAFLOAEKDIADLGALY 442  
 DB 66 SLOWKKTTFEKLQSIARMKQLELHQEGKSYLLTSPSGVAANDTHPT-----SLM 116  
 QY 443 SQRFQKY-KNVEFRSILRLDNADTTGNRTNLVSGSGVLIDPATNTLIVTDTSRVIEK 501  
 DB 117 TSSIKLHFAKAEVVKSL-----TSGGSLVSGSISFERTNLLIQDEPSIQIR 168  
 QY 502 FRKLIDELDPVQAVMTIEARIVEAADGSRDLGVKFG---ATGKKLKNDTSPFGWVNS 558  
 DB 169 IKALVAEMDKPIEQIAIEARIVTDTESLQELGVWGLFOATEQA-----HTIAGSLAAN- 223  
 QY 559 GFGGDDKWAETKINLPI---TAAASISLVRA-ISSGALNLELSASELSKTKTLANPR 614  
 DB 224 GFS-----NIENQUNVFNFSNAPVGSIALQLAKINGRLDLLELALEREKHEIITASPR 278  
 QY 615 VLTQNRKEAKIESGYEIPFTVTSIANGSGSNTTELKXAVLGLTVTPNITPDGQIIMTKVI 674  
 DB 279 LTTNKKASIKQTEIPYVMK--RGKDKSESVEFREAVLGLDVTPI SKDNTILLDLI 336  
 QY 675 NKD---SPAQCASGNQITLCISTKNLNTQAMVNGGTLIVGGIYEDNGNTLTKVPLIGD 731  
 DB 337 TQNTLGAPVVYDKGE--IVSIDKQEIQTQVVAQGETIVLGGVFDHTMTKGVNKPVLIGD 394  
 QY 732 IPVIGNLFTKRGKTDRELLIFITPRIM 760  
 DB 395 LPLLKHVFSQKTERHQKRELIVFTPHII 423

RE:ULT 19  
 Q8XUT8 PRELIMINARY; PRT; 754 AA.  
 AC Q8XUT8;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Putative GSPD-related protein.  
 GN RSP0143 OR RS02977.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OG Plasmid megaplasmid.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Ralstoniaceae; Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]\_TaxID=305;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Ariat M., Billaut A., Brottier P., Camus J.C., Cattolico L.,  
 RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,

RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";  
 RL Nature 415:497-502 (2002).  
 DR EMBL: AL646076; CAD17294.1; -.  
 DR InterPro: IPR001775; Bac GSPD.  
 DR InterPro: IPR004846; GSPII/IIprotein.  
 DR InterPro: IPR005644; NOLW-like.  
 DR InterPro: IPR001440; TPR.  
 DR Pfam: PF00263; GSPII\_III; 1.  
 DR Pfam: PF03958; GSPII\_III; 1.  
 DR PRINTS: PR00811; BCTERIALGSPD.  
 KW Plasmid; Complete proteome.  
 SQ SEQUENCE 754 AA; 81518 MW; B77410CBA02D7D0 CRC64;

Query Match 10.2%; Score 393.5; DB 16; Length 754;  
 Best Local Similarity 23.0%; Pred. No. 6e-14;  
 Matches 141; Conservative 118; Mismatches 234; Indels 121; Gaps 18;

QY 190 PAKQQAASAKQOATAAPAKQOATAAPAKQOATNIDFRKQKNAIGIIEALALGAPAG 249  
 DB 28 FYREGEALSAGKNEASLERFETAAKAOPTNARYAAVLQARDRTVNGWLEAERLRTG 87  
 QY 250 QPDISQOHDHIVTL--KNHTLPTTLQRSLDVAFKTPQVKV--TLKRLNDTQLIITTA 305  
 DB 88 KPDAARKMYERVLLALAPENARARAGLEQGERDRRHANLVQQAEDALKKGERDTALAKLHL 147  
 QY 306 GNKELVNKSAAPGYFTFQVLPKKQNLBSGGVNNAPKFTFG--RKISLDFDQDVEIRTILO 362  
 DB 148 ALAE--NQFRP-----ALALQRIEPTGDSPERALSAAFRKPISEFRDAQMRQVFE 199  
 QY 363 ILAKESGMNIVASDV--NGKMTLSLKDVPMDQALDVMQARNLDMR--OQGNIVNIAPRD 419  
 DB 200 VLSRASGLNFVFDKVRTDQKVTFNRSTVANVVNTVLLTNQLEQRVLDDGNSILIVENT 259  
 QY 420 ELLAKD-----KAFLOAEKDIADLGALYSQNFQLYKNVVEFRSILRLDNADTTGNRT 473  
 DB 260 PAKQDYQDLTVRTFVLSNSDAKV-----ANTV---KT 290  
 QY 474 LVSGSGSVLIDPATNTLIVTDTSRVIEKRLIDELVPAQQVMIEARIVEAADGFSRDL 533  
 DB 291 ILKTR-DIVWDEKRNIVMRDTPDAIQAAEKLAVHMPPEPEVMEVEILEVKRSRLQDL 349  
 QY 534 GYKFGATGKKLKNDTSPAFGWGNSFGGDDKXGAETKINLPITAAANSISLVRAISSGA 593  
 DB 350 GQLPS-----QSLTPLAS-----SGT 369  
 QY 594 LNLE-----LSASELSKTKTLANPRVLTQNRKEAKIESGYEIP-FT 634  
 DB 370 LTLDLHLNTPRIGATVSPLVNVAATDINLANPRVTRNKEKARIQVGRVPNT 429  
 QY 635 VTSIANGSGSNTTELKXAVLGLTVTPNITPDGQIIMTKVINKD-----SPAQCASGNQTI 689  
 DB 430 TTSATGFAENVQVYDVGKLEVEPTVPDNEV--TIRINLEVSDIISQVQTKSGS-IA 486  
 QY 690 LCISTKNLNTQAMVNGGTLIVGGIYEDNGNTLTKVPLIGDIPVIGNLFTKRGKTDPR 749  
 DB 487 YEIGTRNATTLRLKXGQENQILAGLIQEDRVSNGRVFLGDI PALGRFLGQAQSDNKLKS 546  
 QY 750 ELLIFITPRIMGTA 763  
 DB 547 EIVLSITPRIRPA 560

RESULT 20  
 Q8XUS1 PRELIMINARY; PRT; 805 AA.  
 ID Q8XUS1;  
 AC Q8XUS1;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Probable general secretory pathway D transmembrane protein.  
 GN GSPD OR RSC3114 OR RS00567.

|   |   |  |
|---|---|--|
| OS  | Ralstonia solanacearum (Pseudomonas solanacearum).                    |  |
| OC  | Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;        |  |
| OX  | Ralstoniaceae; Ralstonia.   |  |
| NCBI_TaxID=305;   |   |  |
| DB  | [1]   |  |
| RP  | SEQUENCE FROM N.A.  |  |
| RC  | STRAIN=GMI1000;   |  |
| RX  | MEDLINE=21681879; PubMed=11823852;                                    |  |
| RA  | Salaoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,        |  |
| RA  | Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,         |  |
| RA  | Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,      |  |
| RA  | Gaspin C., Lavié M., Moisan A., Robert C., Saurin W., Schiex T.,      |  |
| RA  | Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,              |  |
| RA  | Weissenbach J., Boucher C.A.;   |  |
| RT  | "Genome sequence of the plant pathogen Ralstonia solanacearum.";      |  |
| RL  | Nature 415:497-502(2002).   |  |
| DR  | EMBL; AL646073; CAD16823.1; -   |  |
| DR  | InterPro; IPR001775; Bac_GSPD.  |  |
| DR  | InterPro; IPR004846; GSPII/IiIprotein.                                |  |
| DR  | InterPro; IPR004845; GSPIIproteinC.                                   |  |
| DR  | InterPro; IPR005644; NOLW-like.                                       |  |
| DR  | Pfam; PF00263; GSPII_III; 1.  |  |
| DR  | Pfam; PF03958; GSPII_III_N; 3.  |  |
| DR  | PRINTS; PR00811; BCTRIALGSPD.   |  |
| DR  | PROSITE; PS00875; T2SP_D; 1.  |  |
| KW  | Complete proteome.  |  |
| SQ  | SEQUENCE 805 AA; 82712 MW; FCD75FCAD273416B CRC64;                    |  |
| Query Match 9.9%; Score 380; DB 16; Length 805;                     |   |  |
| Best Local Similarity 24.5%; Pred. No. 3.8e-13;                     |   |  |
| Matches 178; Conservative 106; Mismatches 235; Indels 148; Gaps 29; |   |  |
| QY  | 132 SAFARPAVKAAPA--AP-AKQQGRTYVQVRSIRIQFLY-----PKGTAAAPFTESVV 182     |  |
| DB  |   |  |
| DB  | 30 SLPVTPAFAPPAQASQAPASNGFDEVSLNFVNADLETVWRAVGQATGKFNIVPRVKGTV 89     |  |
| QY  | 183 SVSAPFSPAQAASAKQOTAPAKQOTAPAKQAAPAKQ---TNIDFRKDGKNAGI 239         |  |
| DB  | 90 NL-----VTEKPVTRAQLESLSILRMQGYALVEGNFTKVPEDAKLQGS 137               |  |
| QY  | 240 IELAALGFA--GQPDISQ-----QHDHIVTLK-----NHLPT-TLQSLDVADEFT 284       |  |
| DB  |   |  |
| DB  | 138 PTVSGPGARGGEQVQVFLQVESANLVPVLRPMIAFNITITAYPANNTLVITDYAD 197       |  |
| QY  | 285 PVQKVTILKRLNNDQLIITT-----AGNWELV---NKSAPGYFTFQVLPKKQNLSEGGV 336   |  |
| DB  | 198 NLRRIA-----RIITSDSPAAGETELIALKNVAIDAAATLQKL-----LDPSG- 242        |  |
| QY  | 337 NNAKPTFTGRKISLDFQDVEIRTIQILAKESGMNIVASDS---VNGKMTLSLKDVPWDQ 393   |  |
| DB  | 243 ----TAGGAGAGAALADPSLRTSVVAEPRSNVSVLVEASSAARMAQAKQLAKLDVPGTR 297   |  |
| QY  | 394 ALDL-VMQARNLVRQOGNIVNIAPRDELLAKAKAFLQAE-----KDIA 436              |  |
| DB  | 298 PGNIVVYVFLKN-----ANAVQLATTLRAIVAADATLSASQSGPGQSAQAQAQCPA 351      |  |
| QY  | 437 DLGALYSQNFQI-KYKVV-----EFRSILRLDNADTTGNRNTLVSGRGSVLID 484         |  |
| DB  | 352 TTGTQTQNTQTSYSSSSSGSGSGSGSFRASFGQSLPTTG-----GLIQAD 402            |  |
| QY  | 485 PATNTLIVTTRSVIEFKRLIDELVPAQVMIARIVAAAGDFGRDLGVKF--GATGK 542       |  |
| DB  | 403 PATNALIITASEPYRNLRTVIDDLARRAQVYTESMIVEVTSKASQLGIQWVGAGGP 462      |  |
| QY  | 543 KKLKNDTSAFGWGVS-----GFGGDKWGAETKINLPITAAANSISLVAISS 591           |  |
| DB  | 463 NTYGFQGTNGSGVGNILNLGVIAATVSGSGIGSTAAQTALG---SITGNSVSLNGGNF 519    |  |
| QY  | 592 GAINLE-----LSASSLSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSIANGGS 645      |  |
| DB  | 520 GVFNKNTGLGAILSALSGDSGVNVLSTFNLTIDNEEAKILIQNVPIITGSAQTGSSA 579     |  |
| QY  | 646 NT-----ELKAVLGLVTVPNTIPDQIIMTVKINKDSPAQCASGNQITL----CISTKN 696    |  |
| DB  |   |  |
| DB  | 580 SVTPFTQTFDRKDVGLTLRVKPOITDGMWKM--QIFQESSA-VVNGTQNTATGPTTNVRS 636  |  |
| QY  | 697 LNTQAVENGSTLIYGGIYEDNGNTLTKVPLGLDIPVIGNLFTKRGKTKDRRELLIFT 756     |  |
| DB  | 637 IETNVIANDQGVILVGLGLEDNYQDSEKQVGLGIPVLGALFRSEKSRKTKTNLLVFLR 696    |  |
| QY  | 757 PRIMGTA 763   |  |
| DB  | 697 PYILRTA 703   |  |
| RESULT 21   |   |  |
| Q9WXU3  | PRELIMINARY; PRT; 1285 AA.  |  |
| AC  | Q9WXU3;   |  |
| DT  | 01-NOV-1999 (TrEMBLrel. 12, Created)                                  |  |
| DT  | 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)                     |  |
| DE  | 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)                   |  |
| DE  | COME protein, putative.   |  |
| GN  | TM0088.   |  |
| OS  | Thermotoga maritima.  |  |
| OC  | Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.     |  |
| OX  | NCBI_TaxID=2336;  |  |
| RN  | [1]   |  |
| RP  | SEQUENCE FROM N.A.  |  |
| RC  | STRAIN=MSB8 / DSM 3109;   |  |
| RX  | MEDLINE=99287316; PubMed=10360571;                                    |  |
| RA  | Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,        |  |
| RA  | Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,     |  |
| RA  | McDonald L., Utterback T.R., Malek J.A., Linner K.D., Garrett M.M.,   |  |
| RA  | Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  |  |
| RA  | Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,   |  |
| RA  | Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;                  |  |
| RT  | "Evidence for lateral gene transfer between Archaea and Bacteria from |  |
| RT  | genome sequence of Thermotoga maritima.";                             |  |
| RL  | Nature 399:323-329(1999).   |  |
| DR  | EMBL; AE001695; AAD35182.1; -   |  |
| DR  | TIGR; TM0088; -   |  |
| DR  | InterPro; IPR004846; GSPII/IiIprotein.                                |  |
| DR  | InterPro; IPR001993; Mitoch carrier.                                  |  |
| DR  | Pfam; PF00263; GSPII_III; 1.  |  |
| DR  | PROSITE; PS00215; MITOCH_CARRIER; 1.                                  |  |
| KW  | Complete proteome.  |  |
| SQ  | SEQUENCE 1285 AA; 145209 MW; 057435P921FB0EA5 CRC64;                  |  |
| Query Match 9.4%; Score 363; DB 16; Length 1285;                    |   |  |
| Best Local Similarity 25.0%; Pred. No. 6.9e-12;                     |   |  |
| Matches 105; Conservative 95; Mismatches 152; Indels 68; Gaps 13;   |   |  |
| QY  | 347 KISLDFQDVEIRTIQILAKESGMNIVASDSVNGKMTLSLKDVPWDQALDLMQARN-LD 405    |  |
| DB  | 926 KLTINAEAPLYDLLEEIASELGISVWFVSTPSEKITMKADNVAMEKFDILISQNYGLF 985    |  |
| QY  | 406 MQQGNVNIAPRDELLAKRAFLQAEKDIAADLGALYSQNFOLKYKNVEEFRSILRLDNA 465    |  |
| DB  | 986 DNKNGVYVSKPKQDL-----ARRYIDV---PHNPD-QIKALIEFYG----- 1024          |  |
| QY  | 466 DTTGNRNTLVSGRGSVLIDPATNTLIVTDRSVIEK-PRKLIDELDVPAQVMIARIVE 524     |  |
| DB  | 1025 -----GTVYVDSLNNFMVVTGISETIKRELDNIIEKLKXPTKQIEISAKIVD 1071        |  |
| QY  | 525 AA--DGFGRDLGVKFGATGKKLKNDTSAFGWGVNSGFGDDKMGCAETKINLPITAAANS 582   |  |
| DB  | 1072 RSLDLRSKEITGEL--TGE-----NVNVGSSG-----AEISFSVTDYIDF 1110          |  |
| QY  | 583 ISLVRAISSGAINLSASELSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSIANGG 642     |  |
| DB  | 1111 EKIFGEILLNTLSLQFSDQKNTLDDILASPRIVTTSKGEARILIGDRIPYVDT--NGD 1168  |  |
| QY  | 643 STNTTELKAVLGLVTVPNTIPDQIIMTVKINKDSPAQCASGN--QTILCISTKNLNTQ 700    |  |
| DB  | 1169 GTPEVQFLETGIELSITPFVRSDDTIEDLDFVKASEP-----GNVINEVPGERTREATH 1223 |  |

701 AMVNGGTLVGGIYEDNGNTLTQVPLGDIPIVIGNLFKTRGKTDRELLIFITPRIM 760  
1224 LIVXNGSTIIIGGLIREVTNVTESKLPFLGDLPIVGQFFRTKSENKEKRDVIFLIVRVV 1283

RESULT 22  
Q8VRNO PRELIMINARY; PRT; 616 AA.  
AC Q8VRNO;  
DT 01-MAR-2002 (TREMELrel. 20, Created)  
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)  
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
DE Hypothetical type II secretion protein GspD.  
GN GSPD.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=83/39;  
RA Tauschek M., Strugnell R.J., Strugnell R.A., Robins-Browne R.M.;  
RT "Identification of a type II protein secretory pathway required for  
RT the secretion of heat-labile enterotoxin by enterotoxigenic  
RT Escherichia coli."  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF426313; AAL60184.1; -  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSPDII/IIprotein.  
DR InterPro; IPR005644; NOLW-like.  
DR Pfam; PF00263; GSPDII\_III; 1.  
DR Pfam; PF03958; GSPDII\_III; 3.  
DR PRINTS; PRO0811; BCTERIALGSPD.  
DR SEQUENCE 616 AA; 66319 MW; 7069455A3F19A8B4 CRC64;

Query Match 9.4%; Score 362.5; DB 2; Length 616;  
Best Local Similarity 24.3%; Pred. No. 2.5e-12;  
Matches 131; Conservative 96; Mismatches 187; Indels 125; Gaps 19;

QY 285 PVQKVTLEKLNNDTQIIITTAGNWLNVKSA--PGYF 320  
DB 79 PVRNVSURELAPILQWIDSAGSNVNYDPSNVIMLTGRASVVERTEVIQVRDHAGNR 138  
QY 321 TFQVLPKQNLSEGVNNAKPTFTGKISLDQPDVEIRTIQILAKESGMN-----I 372  
DB 139 TEEVIP-----LDNASAS-----EIARVLESLTNSGENQATLKSQI 176  
QY 373 VASDSVNGKMTLSLKQVDPWDQALDVMQARNLD--MRQOGNIVNIAPRDELLAKDAFLQ 430  
DB 177 VADERTN-SVIVSGDPATRDQRRLLI--RRLDSEMERSGN-----213  
QY 431 AEKDIADIGALYSQNFQKYNVEEFSIRLDNADTTGNR---NTLVSGRG--SVLID 484  
DB 214 -----SQVFYLYKSKAEDLDVLRKQVSGTTLTAKEEAEAGTVGSGREIVSIAAS 261  
QY 485 PATNTLIVTDTRSVIEKFKLIDELDPVPAQVMIEARIVEAADGFSRDLGVKFGA--TGK 542  
DB 262 KHSNALIVTAPQDINQSLQSVIEQLDIRRAQVHVEALIVEAEGSNINFGVQWASKDAGL 321  
QY 543 KKLKNDT-----SAFGWGN-----SGFGDDKWAET---KINLPITAAANSISLVRAIS 590  
DB 322 MQFANGTQIPIGTLGAATSOAKPKQGSTVISENGATTINPDNGDLSTLAQLLGSFGTA 381  
QY 591 SGALNLE-----LSASELSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANGGSSTN 646  
DB 382 VGVKGDMMALVQAVKNDSSNVLTSTPSTITLDNQEAFFWQGVDP-VLTGSTVGSNNNSN 440  
QY 647 ----TELKAVLGLVTNPITPDGQIMTVKINKDSPAQCASGNQTI-LCISTKNLNTQA 701  
DB 441 PFNTVERKKGIMLVKVTPOINEGNVQVMI----EQEVSKVEGQTSLDVVFGERKLKTV 496  
QY 702 MVENGGLTIVGGIYEDNGNTLTQVPLGDIPIVIGNLFKTRGKTDRELLIFITPRIM 760

497 LANDGELIVGLGMLDDQAGSEVAKVPLLDIPLIGNLFKSTADKKEKNLMVFIPTIL 555

RESULT 23  
Q8VPC8 PRELIMINARY; PRT; 616 AA.  
AC Q8VPC8;  
DT 01-MAR-2002 (TREMELrel. 20, Created)  
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)  
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
DE Hypothetical type II secretion protein.  
GN GSPD.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=H10407;  
RA Tauschek M., Strugnell R.A., Robins-Browne R.M.;  
RT "Identification of a type II secretory pathway required for secretion  
RT of heat-labile enterotoxin by Enterotoxigenic Escherichia coli."  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY056599; AAL10693.1; -  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSPDII/IIprotein.  
DR InterPro; IPR005644; NOLW-like.  
DR Pfam; PF00263; GSPDII\_III; 1.  
DR Pfam; PF03958; GSPDII\_III; 3.  
DR PRINTS; PRO0811; BCTERIALGSPD.  
DR SEQUENCE 616 AA; 66353 MW; 7069455A3F19A654 CRC64;

Query Match 9.4%; Score 362.5; DB 2; Length 616;  
Best Local Similarity 24.3%; Pred. No. 2.5e-12;  
Matches 131; Conservative 96; Mismatches 187; Indels 125; Gaps 19;

QY 285 PVQKVTLEKLNNDTQIIITTAGNWLNVKSA--PGYF 320  
DB 79 PVRNVSURELAPILQWIDSAGSNVNYDPSNVIMLTGRASVVERTEVIQVRDHAGNR 138  
QY 321 TFQVLPKQNLSEGVNNAKPTFTGKISLDQPDVEIRTIQILAKESGMN-----I 372  
DB 139 TEEVIP-----LDNASAS-----EIARVLESLTNSGENQATLKSQI 176  
QY 373 VASDSVNGKMTLSLKQVDPWDQALDVMQARNLD--MRQOGNIVNIAPRDELLAKDAFLQ 430  
DB 177 VADERTN-SVIVSGDPATRDQRRLLI--RRLDSEMERSGN-----213  
QY 431 AEKDIADIGALYSQNFQKYNVEEFSIRLDNADTTGNR---NTLVSGRG--SVLID 484  
DB 214 -----SQVFYLYKSKAEDLDVLRKQVSGTTLTAKEEAEAGTVGSGREIVSIAAS 261  
QY 485 PATNTLIVTDTRSVIEKFKLIDELDPVPAQVMIEARIVEAADGFSRDLGVKFGA--TGK 542  
DB 262 KHSNALIVTAPQDINQSLQSVIEQLDIRRAQVHVEALIVEAEGSNINFGVQWASKDAGL 321  
QY 543 KKLKNDT-----SAFGWGN-----SGFGDDKWAET---KINLPITAAANSISLVRAIS 590  
DB 322 MQFANGTQIPIGTLGAATSOAKPKQGSTVISENGATTINPDNGDLSTLAQLLGSFGTA 381  
QY 591 SGALNLE-----LSASELSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANGGSSTN 646  
DB 382 VGVKGDMMALVQAVKNDSSNVLTSTPSTITLDNQEAFFWQGVDP-VLTGSTVGSNNNSN 440  
QY 647 ----TELKAVLGLVTNPITPDGQIMTVKINKDSPAQCASGNQTI-LCISTKNLNTQA 701  
DB 441 PFNTVERKKGIMLVKVTPOINEGNVQVMI----EQEVSKVEGQTSLDVVFGERKLKTV 496  
QY 702 MVENGGLTIVGGIYEDNGNTLTQVPLGDIPIVIGNLFKTRGKTDRELLIFITPRIM 760  
497 LANDGELIVGLGMLDDQAGSEVAKVPLLDIPLIGNLFKSTADKKEKNLMVFIPTIL 555

|   |  |   |              |
|---|--|---|--------------|
| Db  | 400  | DEVSMKINLEVSSLTNQITTRSGTVAYQIGTRANTVYKLUHDESTQALLAGLTKTQOTSSA       | 459          |
| QY  | 724  | TKVPLGLDIPVIGNLFTKRGKKTDRELLIFITPRIM                                | 760          |
| Db  | 460  | ARIPGLGDIPLGLRFLSSQTDNGVRNEIVLSITPRVV                               | 496          |
| RESULT 25   |  |   |              |
| Q8PFJ1  |  |   |              |
| ID  | Q8PFJ1   | PRELIMINARY:  | PR7; 692 AA. |
| AC  | Q8PFJ1   |   |              |
| DT  | 01-OCT-2002  | (TREMBlrel. 22, Created)  |              |
| DT  | 01-OCT-2002  | (TREMBlrel. 22, Last sequence update)                               |              |
| DT  | 01-MAR-2003  | (TREMBlrel. 23, Last annotation update)                             |              |
| DE  | Type II secretion system protein D.                                    |   |              |
| DE  | XCSO OR XAC0695.   |   |              |
| GN  | Xanthomonas axonopodis (pv. citri).                                    |   |              |
| OS  | Xanthomonas axonopodis (pv. citri).                                    |   |              |
| OC  | Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;        |   |              |
| OC  | Xanthomonadaceae; Xanthomonas.   |   |              |
| OX  | NCBI_TaxID=92829;  |   |              |
| RN  | [1] _TaxID=92829;  |   |              |
| RP  | SEQUENCE FROM N.A.   |   |              |
| RC  | STRAIN=J06 / ATCC 13902 / XV 101;                                      |   |              |
| RX  | MEDLINE=2202145; PubMed=12024217;                                      |   |              |
| RA  | da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,    |   |              |
| RA  | Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,   |   |              |
| RA  | Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,          |   |              |
| RA  | Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,     |   |              |
| RA  | Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,     |   |              |
| RA  | Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,            |   |              |
| RA  | Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,                |   |              |
| RA  | Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,    |   |              |
| RA  | Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,      |   |              |
| RA  | Martins E.C., Meidanis J., Menck C.P.M., Miyaki C.Y., Moon D.H.,       |   |              |
| RA  | Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,   |   |              |
| RA  | Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.P.,          |   |              |
| RA  | Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D., |   |              |
| RA  | Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,              |   |              |
| RA  | Setubal J.C., Kitajima J.P.;   |   |              |
| RT  | "Comparison of the genomes of two Xanthomonas pathogens with differing |   |              |
| RT  | host specificities."   |   |              |
| RL  | Nature 412:459-463(2002).  |   |              |
| DR  | EMBL; AE011699; AAM5584.1; -.  |   |              |
| DR  | InterPro; IP001775; Bac GSPD.  |   |              |
| DR  | InterPro; IP004846; GSPII/IIIProtein.                                  |   |              |
| DR  | InterPro; IP005644; NolW-like.   |   |              |
| DR  | Pfam; PF00263; GSPII III; 1.   |   |              |
| DR  | Pfam; PF03958; GSPII-III N; 3.   |   |              |
| DR  | PRINTS; PR00811; BCTERIALGSPD.   |   |              |
| KW  | Complete proteome.   |   |              |
| SQ  | SEQUENCE 692 AA; 72938 MW; ACF1A0D46FCFB62 CRC64;                      |   |              |
| Query Match   | 9.2%; Score 352.5; DB 16; Length 692;                                  |   |              |
| Best Local Similarity   | 22.3%; Pred. No. 1 le-11;  |   |              |
| Matches 149; Conservative 124; Mismatches 249; Indels 147; Gaps 23; |  |   |              |
| QY  | 198  | SAKQOTAPAKQOTAPAKQQAAPA-KOTNIDFRKDGKNGAGHIEAALGFAGQPDISOQ           | 256          |
| Db  | 11   | SATLLALPAVPMPTALHA---ADAPAVRLQDVLR-----AFIQDVS- 50                  |              |
| QY  | 257  | HDHIIIVLKHNHTPTTLQSLDVADPKTPQKVTKRLNNDTQIITTAGNWLNVKSA 316          |              |
| Db  | 51   | -----ATGTTIVDTRVQGSNNVARAQAMSEADLLGML-----LAVLRANG--LIAVSSG 98      |              |
| QY  | 317  | PGYFTFQVLP-----KKQNLSEGGVNNAPKFTTGRKISLDQDVVEIRTIQL 364             |              |
| Db  | 99   | PS--TYRVIDPDTAAQPGSAAGNLGFATQVFT-----LQRVDSAAEILKPLGR 149           |              |
| QY  | 365  | -----AKESGMNIVASDVNG--KMTLSLKDVDPWDQALDVMQARNLDM- 407               |              |
| Db  | 150  | GGVIMAMPQGNLSIADYADNLRIRTLVAQIDTDRAAIDTTLTNSAACEIARTLSLF 209        |              |
| QY  | 408  | -----QQGNIVNIAPRDE-----LLAKDKAFIQC-AEKDIADLGALYSON-----FOLKYKNV 453 |              |

```

Db 210 GCGERSNVLSVLPVSSNSLIVRGDPALVORVVRVAVDLDGRAERRGVSVVRLOHASA 269
Qy 454 EEFRSILR-----LDNADITGNRNTLV-----SGKGSVLID-PAT 487
Db 270 EQLLPVLOQLVGTGPNGEAQVGDTRLATIDVAASGAATQVIAAAGKRPVIVKYPGS 329
Qy 488 NTLIVDTTRSVEIKFRKLIDELDPVPAQVMIEARIVEAADGSRDLGVKFGATGKKLKN 547
Db 330 NALLIINADPETORALMDVIRQOLVHREQVLVEAIVVEISDRAKRLGVQLLAG-----RN 385
Qy 548 DT-----SAFGWGVNSGFGGDD-----KWGAETKINLPITAAANS 582
Db 386 GTVPLVATVSYGASPGIVPLAAAGATRSNADDDSDVLEQARNVAAQSLGL-----SGGL 441
Qy 583 ISLVRAISSGALNLELSASELSKTKTLANPRVLTQNRKEAKIESGYEIPFT---VTSIA 539
Db 442 IGLAGOSNDAVFGMIIDAVKSGDGSNLLSTPSIMTLDNEQARILVGEVPIITGVELGAA 501
Qy 640 NGGSSTNTEKXAVLGITVTPNITPDQIIMTVKINKDSPAQCASGNQITLCISTKNLNT 699
Db 502 NDNPFRIOQDVGVEVLRPQINTAGGITLAKQEVSAIAGVPSAQSSSELVFNKRQIET 561
Qy 700 QAMVNGGTLIVGGIYEDNGNTLTKVPLLGDPVIGNLFPKTRGKKTDRRELLIFITPRI 759
Db 562 RVVVGNAIVAGLGLDNDQNDQVKEVPLGDLGVGLGALFRHKSNRDKTNLMVFIPTI 621
Qy 760 MGTAGNSLR 768
Db 622 IRDAADAQR 630

RESULT 26
O80264 PRELIMINARY; PRT; 500 AA.
AC O80264; 1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to gene IV protein: Acc# A04268.
OS Vibrio cholerae filamentous bacteriophage fs-2.
OC Viruses; ssDNA viruses; Inoviridae; Inovirus.
CX NCBI_TaxId=83201;
RN [1]
RP SEQUENCE FROM N.A.
RA Ikema M.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98361036; PubMed=9695923;
RA Ikema M., Honma Y.;
RT "A novel filamentous phage, fs-2, of Vibrio cholerae O139.";
EL Microbiology 144:1901-1906(1998).
DR EMBL; AB002632; BAA33484.1;
DR InterPro; IPR001775; Bac_GSPD.
DR InterPro; IPR004846; GSP11/IIIProtein.
DR Pfam; PF00263; GSP11_III; 1.
DR PRINTS; PR00811; BCTERIALGSPD.
DR PROSITE; PS00875; T2SP D; 1.
SQ SEQUENCE 500 AA; 54021 MW; 402B69CEBACAFB30 CRC64;

Query Match 9.0%; Score 347; DB 9; Length 500;
Best Local Similarity 26.5%; Pred. No. 1.4e-11;
Matches 126; Conservative 92; Mismatches 182; Indels 76; Gaps 16;

Qy 307 NWELVNSAAPGYFTFVLPK--KONLESQGVNNAKPTFT---GRKISL-----350
Db 11 NEKVN--PAPRWIFLFEKIRRGATSG--HAAKQSTILSAQQLLILMLSLPAAH 65
Qy 351 ---DFQDVEIRTIQILAKSGGMNIVASDVNGKMTLSLKD-VP--WDQALDVMQARNL 404
Db 66 APFESSDTPIAEFASWYSQGTGKIVLGQGVLSVSTAPDLVPAEYPAPFDSVLRAGY 125

```

```

Qy 405 DMRQGN--IVNIAPRDELLAKDKAFLOAEKDIAAD-IGALYSQNFQFKYKNVEEFSILR 461
Db 126 YLVKDGNAIVYKIAAP-----EAEVITPAIVKLYRFNRIIRNSKUSDLVQSTLK 173
Qy 462 LDNADITGNRNTLVSGRGSVLIDPATNTLIVDTTRSVEIKFRKLIDELDPVPAQVMIEAR 521
Db 174 ATSEFEVKDKQV--DNYSVEILPNTNALIVSGTAQOLEKLDVLLSAIDVFPQRIFEAV 230
Qy 522 IVEAADGSRDLGVKFGATGKKLKNDTSAFGWGVNSGFGGDDKMGAEKTNLPITAAAN 581
Db 231 ITETELGNSLGNLQAADF-----AGFVTNLVNASKKNLFIPIESGD 276
Qy 582 SISLVRAISSGALNLELSASELSKTKTLANPRVLTQNRKEAKIESGYEIPFTVT-SIAN 640
Db 277 FNALVKALSGS-----SDTRLLSRPNLIMDRRGYITVGNVFLVSNSTTD 324
Qy 641 GGSST-NTTELKXAVLGITVTPNITPDQIIMTVKINKDSPAQCASGNQITLCISTKNLNT 699
Db 325 GGTSVQRIERKDVGSLEVTTPVMGDDVILV---INQESSVTDSTIAADIITNKRLMT 381
Qy 700 QAMVNGGTLIVGGIYEDNGNTLTKVPLLGDPVIGNLFPKTRGKKTDRRELLIFI 755
Db 382 TVAVKSGGTIVLGGILSDERKNVESGVPLKDTPLIGLFRSTSTKNVQKELRVVI 437

RESULT 27
Q92FY0 PRELIMINARY; PRT; 649 AA.
AC Q92FY0; 1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane secretion protein Q.
CG XCPQ.
OS Pseudomonas alcaligenes.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
CX NCBI_TaxId=43263;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=M-1;
RX MEDLINE=99040991; PubMed=9823657;
RA Geritise G., Ure R., Bizouillier F., Quax W.J.;
RT "The phenotype enhancement method identifies the Xcp outer membrane
RT secretion machinery from Pseudomonas alcaligenes as a bottleneck for
RT lipase production.";
RL J. Biotechnol. 64:23-38(1998).
DR EMBL; AF092918; AAC8352.1;
DR InterPro; IPR001775; Bac_GSPD.
DR InterPro; IPR004846; GSP11/IIIProtein.
DR InterPro; IPR004845; GSP11proteinc.
DR Pfam; PF00263; GSP11_III; 1.
DR Pfam; PF03958; GSP11_III; 3.
DR PRINTS; PR00811; BCTERIALGSPD.
DR PROSITE; PS00875; T2SP D; 1.
SQ SEQUENCE 649 AA; 68612 MW; 2DE50A042C19E684 CRC64;

Query Match 8.9%; Score 341.5; DB 2; Length 649;
Best Local Similarity 22.4%; Pred. No. 4.1e-11;
Matches 136; Conservative 94; Mismatches 183; Indels 195; Gaps 18;

Qy 315 AAPGYFTFVLPKKNLESQGVNNAKPTFTGKISLDFQDVEIRTIQILAKSGMNIVA 374
Db 24 AAP-----LPLVHAAPFVAVSQGAETWT-----INMKDADIRDFIDVQAISGETFV 71
Qy 375 SDSVNGKVT-----LSLKDYVPWQALDVMQARNLDMRQGNIVNIAPRDELLAKOKAF 428
Db 72 DFRVKGQVTVISKTPPLGLEEV--YQLFVSVNMTHTGFSVLAQGDQARIVPVTEARSGANS 129
Qy 429 LOAEXDI-----ADLGALYSQNFQ---KYKNVEEFSR 459

```



Db 130 RSAPDDVQTELIQVHTSVNELIPRLVLPQNGHLAAVAASNALIISDRANIEREL 189  
Qy 460 LR-----LONADTTGN-----RNTLVSGRGVLDIPATNTLIV----- 492  
Db 190 IAEIDAQGGDYNVINLQHWLWDAEALNNAVNREKNKSAGTRVIAADARTNRLILGPP 249  
Qy 493 -----TDRSVIE----- 500  
Db 250 AARQELANLARSLDIPSTRSANARVIRLRHSDAKSLAETLCDISEGLKTAEGGGEAAASK 309  
Qy 501 -----KFKLIDELDVPAQQMIEARIVEAAGSRLGV 535  
Db 310 PONTILIRADESLNALVLADPDVTATLEEIVRNLDVPAQVMVEAAIVE:SGDISDALGV 369  
Qy 536 KFGATGKKLKNDRSAFGWGVNSGFGG-----DDKGAETKINLPITAA 579  
Db 370 QWAVDA-----RGGTGGLG-GVNFNTGLSVCTVLKATQNEEIPD-----LTLP 413  
Qy 580 ANSISLVRASIGALNELSASELSKTKTLANPRVTQNRKEAKIESGVEIIPF---TVT 636  
Db 414 DGAIIIGIGTENFGALITALSAN---SKSNLLSTPSLLTLONQEAELVQNVPPOTGSYT 470  
Qy 637 SIANGSGSTNTEKKAVALGLT--VTPNITPDQIIMTV--KINKDSPAQACASGNQTLICI 692  
Db 471 TDASGANNPFTIEREDIGVTLKVTPHINDGATLREVEQEISIAPSAGVNAQVDLVT 530  
Qy 693 STKNLTQAMVENGTLIVGIYBEDNGNTLTKVPLLDIPVIGNLFTKRGKKTDRRELL 752  
Db 531 NKRSIKSVILADGQVIVLGLIQDDVTSTDSKVPLLDIPLIGLFRSTKDTWKVKNL 590  
Qy 753 IFITPRIM 760  
Db 591 VFLRPTIV 598

RESULT 28  
Q988A3 PRELIMINARY; PRT; 708 AA.  
AC Q988A3;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DE General secretion protein D.  
GN MLL6829.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kureko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
Mesorhizobium loti";  
RL DNA Res. 7:331-338(2000).  
DR EMBL; AP003010; BAE53047.1; --  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSP1/IIIProtein.  
DR InterPro; IPR005644; NOLW-like.  
DR Pfam; PF00263; GSP11\_III; 1.  
DR Pfam; PF03958; GSP11\_III\_N; 3.  
DR PRINTS; PR00811; BCTERIALGSPD.  
KW Complete proteome.  
SQ SEQUENCE 708 AA; 75560 MW; CSD991A03DF9486 CRC64;

Query Match 8.8%; Score 338.5; DB 16; Length 708;  
Best Local Similarity 21.6%; Pred. No. 6.9e-11;  
Matches 156; Conservative 137; Mismatches 265; Indels 163; Gaps 30;

Qy 147 AKOQGRCTVYQVRSIRIOTLYFGKTTAAAPFTESVVSAPSPAKQQAASAKQOATAAP 206  
Db 3 SKPSHCVLFTLLAVAGCTSAFGKDF-----FTTIDSLHAKNSPLRAGYSGPAATSSAS 58  
Qy 207 AKQOTAAPAKQ-----QAAAPAKQTNID-----FRKDGKNAGIIEI--AALGFAGQPDIS 254  
Db 59 AARFNGAQYQGTGQFVSSGAPVTKVTSQSGKFEINLVNPIADAAKAVLGD----- 112  
Qy 255 QOHDIHIVTLKHNTLPTTLQRLSLOVADFKTPVQKVTLLKRLNNDTQIIITAGNWLNVKS 314  
Db 113 -LHLNIVDPVVOG-TVTLTQTS-----QPVSDALVDI-LQSALAVNAAG-----ITSR 158  
Qy 315 AAPGYTFFVLKPKNLESGGVNAPKTF--TGRKIS-LDFQDV--EIRTILO----- 362  
Db 159 AG-----TYQIVLSEIMASTPVPSTSPSGFGVKVQVLQFTAADEMKIILEPITQ 214  
Qy 363 --ILAKESGNIV--ASDS-VNG-KMTLSLKDVPMDQALDLMQ-----ARNLD 405  
Db 215 GSVLRVDSTRNITVAGSDNLNAREAVSDFVDMWGMGMSVALHPLKTSKPEAAVAEL 274  
Qy 406 M---ROOG--NIVNIAPDEL-----LAKDKAFI-QAEKDIADLGALYSON-----FQ 447  
Db 275 SIFGTKEGPGAKLIQIPNDRLSVLVITSRPAYLARAATWINKLORLAETNESQLFVYQ 334  
Qy 448 LKYKNVEEFSILR-----LDNADTTG-----N 470  
Db 335 IQNRPAKELASVLSVLGTTVKTGEGSGSNVAPQDTPIAMQSDGVTAPLTGSPSLPQ 394  
Qy 471 RNTLVSGRSLVDPATNTLIVTDRSVIEKFRKLIDELVDPAQVMIEARIVEAAGDFS 530  
Db 395 QNQAPAHATVADVAVENNALLQTTARDYQRIEQLSKVDVLTQVMLEAVAEVT----- 450  
Qy 531 RDLGVKFGATGKKLKNDRSASF--WGVNSGFGGDDKMGAEKINLPITAAANSISLVR- 587  
Db 451 -----LNDDLKYGLRWFENG-----GTKSVTVDAKAAAATLPG 486  
Qy 588 ---AISGALNELSASELSKTKTLANPRVTQNRKEAKIESGVEIPTVTTSIANGSS 644  
Db 487 FNMYSATENIQVTNALSKITDVNVIAPTIMALNNQKAILQVGDQVPLTQOSQDTNG 546  
Qy 645 T-----NTELKKAVALGLTVPNITPDQIIMTVKINKDSPACASGNQILCISTKNLAT 699  
Db 547 SAPIINSVMKDTGVILTVTPRNAGRVMLDIQEVSVNTKTDSSDIDSPITQKQVOT 606  
Qy 700 QAMVENGTLIVGVIYEDNGNTLTKVPLLDIPVIGNLFTKRGKKTDRRELIIFITPRI 759  
Db 607 RVLVNDGESLALGLIQNNVSDRSQVPLGDIPILGNAPKQKDDTIRRTETELIIFIRPHV 666  
Qy 760 M 760  
Db 667 V 667

RESULT 29  
Q8PGT2 PRELIMINARY; PRT; 763 AA.  
ID Q8PGT2  
AC Q8PGT2;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE General secretion pathway protein D.  
GN XPSD OR XAC3534.  
OS Xanthomonas axonopodias (pv. citri).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xanthomonas.  
OX NCBI\_TaxID=92829;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=306 / ATCC 13902 / XV 101;  
RX MEDLINE=2022145; PubMed=12024217;  
RA da Silva A.C.R., Perro J.A., Reinach P.C., Farah C.S., Furlan L.R.,  
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,

|    |  |
|----|--|
| RA | Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo I.E.A.,          |
| RA | Camarotte G., Cannavan F., Cardoso J., Chambergó P., Ciapina L.P.,     |
| RA | Ciacelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,      |
| RA | Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,            |
| RA | Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.T.,              |
| RA | Katsuyama A.C., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,    |
| RA | Locati A.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,      |
| RA | Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,       |
| RA | Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,   |
| RA | Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.P.,          |
| RA | Spinola L.S.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D., |
| RA | Trindade dos Santos M., Truffi D., Tsai S.M., White F.C.,              |
| RA | Setubal J.C., Kitaajima J.P.;  |
| RT | "Comparison of the genomes of two Xanthomonas pathogens with differing |
| RT | host specificities.";  |
| RL | Nature 417:459-463(2002).  |
| DR | EMBL; AB012002; AA38377.1; --  |
| DR | InterPro; IPRO01775; Bac_GSPD.   |
| DR | InterPro; IPRO04846; GSPII/Ilioprotein.                                |
| DR | InterPro; IPRO04845; GSPIIpoteinc.                                     |
| DR | InterPro; IPRO05644; NolW-like.  |
| DR | InterPro; IPRO03522; SecIII_OMPg.                                      |
| DR | InterPro; IPRO00897; SRP54.  |
| DR | pfam; PF00263; GSPII_III; 1.   |
| DR | pfam; PF03958; GSPII_III_N; 3.   |
| DR | PRINTS; PR00811; BCTERIALGSPD.   |
| DR | PRINTS; PR01377; TYPE3OMGPROT.   |
| DR | PROSITE; PS00300; SRP54; 1.  |
| DR | PROSITE; PS00875; T2SP_D; 1.   |
| KW | Complete proteome.   |
| SQ | SEQUENCE 763 AA; 79494 MW; 5EE2BF0B2C662667 CRC64;                     |
|    |  |
|    | Query Match 8.8%; Score 337; DB 16; Length 763;                        |
|    | Best Local Similarity 20.2%; Pred No. 9,3e-11;                         |
|    | Matches 156; Conservative 113; Mismatches 268; Indels 234; Gaps 27     |
| Qy | 165 TLVPGKTTAAAPFTESVSUSAPSAPAKQAASAKOOTAAPAKQ-----OTAAPAKQ 217        |
| Dd | 24 TTPPPDVRENALDFQVGAAGATRPAPQRADGDANAKSPVIRRGSGTMINQOASAPA 83         |
| Qy | 218 QAAAPAKQNIDFRKGKNAGIIELAALG-PAGOPDISQQHDHIIITLKHHTLPPTTLORS 276    |
| Dd | 84 PTLGMASSGSATFNPEGESLQAVVKAILGLMLGQ-----NYVIAPCVQGT 128              |
| Qy | 277 LVVADPKTVPQVKTKLRNLNDTQLIIITAGNWELVNKSAPGVFTQVLPKXNQIESGV 336      |
| Dd | 129 VTLLA--TENPVSPAQLN----LLEVLGMNNARWVFSGG--RNVIVPADQAL-AGTV 177      |
| Qy | 337 NNAPKT-----FTGRKISLDFOV-EIRTILQ-----ILAKESGNMIVASDSVNG 380         |
| Dd | 178 --APSTASPASAARGEVRVVPVKFISASEMKVLEPYARNVAIVGTDPARNVITLCGTPA 235    |
| Qy | 381 KM-----TSLSKDVPWDQALDLVM-----QARNLDM 406                           |
| Dd | 236 ELENYLRTVCIEFDVMDLGMVGVPFIQSOGKAKEVSADLEKFVEGOSKTPSAGMPFMPL 295    |
| Qy | 407 ROQGNTVINTPADELLAKDKAPLOAEKDIDLGYLSQNFQLKYXNVEFRSIL----- 460       |
| Dd | 296 ENANAVLIVITPOPRYLDQIQOWLDRIDSDAGGGVRLFS--VELYKIKAKOLLAPRLSEVFG 353 |
| Qy | 461 RLNDAD-----TTGNRNVTIS----- 476                                     |
| Dd | 354 RNSGGDSNASLAGSETSVLGGTLGNRRDSLSSGSGMTGSGIGSDGSGSSGSSFGSGS 413      |
| Qy | 477 -----GRGSVLIDP-----ATNTLIIVDTRSVIERFKILI 506                       |
| Dd | 414 GSSSGGLNGSLQLSPRSNGNGAVTLEVAGDKVGSVAETNLTLLVRSITPQAWSSIRDVI 473    |
| Qy | 507 DELDVDPAQOVMEEARIVEAADGFSDLVGKCATGKKLKNDTSAFGWNGSVSGFGDDKW 566     |
| Dd | 474 EKULDVPMQMVHEAQVAEV-----NLTCALSIGNV-----W 504                      |
| Qy | 567 GAETKINLRI-----TAAANSTI--SLVRAISSGAINLE-----LGASESL 604            |

Db 505 YENAVTTTFPNADSGGPAFPSAGRNWGDIAGRVTDGGVATFLGKNAALISALDVT 564

Qy 605 SKTKTLANPRVLVLTQNRKAKIESYEIPFTVTSIANG----GSSTNELKKAVALGLTVP 660

Db 565 SNVRLLQTPSVFVRNNAEATLVNGSRIPINSTINTGLGSDSYSSVQVIDTGVLTKVRP 624

Qy 661 NITPDG---GIIMTVKINKDSPACASGNQTL----C---ISTKNLNTQAMVENGGL 709

Db 625 RYTKDGMVFLDVIQEIISTPGARPAACATAATTVNSAACNVDIRRVKTEAAVQSGDM 684

Qy 710 IVGGYIEEDNGNTLTKVPLGLDIPVIGNLTKRKTKTDRRELLIFTPRIM 760

Db 685 MLAGLIDDDTTNGSVGVFLSKLEIVGALFGRKTQNTDRREVLTIPSIV 735

RESULT 30

Q8RTI3

ID Q8RTI3 PRELIMINARY; PRT; 689 AA.

AC Q8RTI3;

DT 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

WmpD.

OS Pseudoalteromonas tunicata.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;

OC Alteromonadaceae; Pseudoalteromonas.

OX NCBI\_Taxid=87626;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=D2;

RA Egan S.; Kjelleberg S.;

RA "Correlation between pigmentation and antifouling compounds produced

RT by Pseudoalteromonas tunicata";

RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF412448; AAL76242.1; --

DR InterPro; IPR004846; GSP11/IIIProtein.

DR InterPro; IPR005644; NoIw-like.

DR Pfam; PF00263; GSP11\_III; 1.

DR Pfam; PF03958; GSP11\_III\_N; 3.

SQ SEQUENCE 689 AA; 75105 MW; 5C6A894CFB2761A5 CRC64;

Query Match 8.7%; Score 336.5; DB 2; Length 689;

Best Local Similarity 21.4%; Pred. No. 8.6e-11;

Matches 129; Conservative 115; Mismatches 212; Indels 147; Gaps 22

Qy 216 KQAAAPAKQTNI----DFRKDG-----KNAGIIEALAL--GFAGQPD---ISQ 255

Db 106 KYKSSDAAKSNVPVLGDDFDVQGDMLVTRVVRVKNVSVQELGPIIRQFSQDKDGGHVN 165

Qy 256 QEDHIIVTLKHTLPTTLQRSLDVADF--KTPQKVTLKELNDTQLIITAGNWLNVK 313

Db 166 YNPSNVLMWTGHA---SSVNRLEVIIRLVDAQGQQQVDIVKLRVATSADVSVV-- 215

Qy 314 SAAPGVFTFQVLPKPKQNLSEGVNAPKTTGTGRKISLDFODVEIRITLOILAKESGNIV 373

Db 216 -----VDNIYKPASGK-----SDIAPFLIPKVADERTNSVI 247

Qy 374 ASDSVNGKMTLSLKDVPDQALDLVQARNLDMRQGN-----IVNTPRDELLAKDKAPL 429

Db 248 VSGEAQAR-----ERAITIILKLDLDD-ELETQNTKVFYINAKAEDLV---KVLUQ 293

Qy 430 QAEKDIADLGALYSQFOLKYKNVEEPRFIRLNDADTTGNRNTLVSGRGSVLID--PAT 487

Db 294 GYSKTIAE-----EQKGAKTSSRRGNDISIEAHPNS 325

Qy 488 NTLIVTDRSVIEKRFKILDELDPVPAQVMIEARIVEAADGFSRDLGVF----- 537

Db 326 NSLIVTAQPDINRSLEGVIAKLDVRRQAQVLVEAIIVEVFEGDGVNLFQWINKQGMQLQF 385

Qy 538 -----CATGKKLKNDTSAFGVGNVSFGDDKRWGAETKLNLPITAAANSISLV 586

Db 386 NNGTTPVSGLSVAGELARDTKTKTVLGNNEGASANOYE---ETKEG-DLTALASLLGV 441

QY 587 RAISSGALNLE-----LSASESLSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSIANGG 642  
DB 442 NGLALFARGDGAALQAVSTDTNSILATPSVTTWDNEEASMIYQVEVPI-ITGSQTGN 500  
QY 643 SSTN-----TELKAVLGLTVTNIPTDGOIIMTVKINKSPQAQCSAGNQT-ICISTKNL 697  
DB 501 NNTNPFQTVRQGVKIKVATQINDGSVQLTI---EQEVSSVSGATAVDITINKREV 556  
QY 698 NTOAMVENGGLTVIGIYEEDNGNTLTKVPLIGDIPVIGNLFTKTKGKTDRELLIFITP 757  
DB 557 TTVTLADGAMVVLGLDIEDVQESVSKVELLGDUPIGHLPFS\*STNRKRKNLLIFIRP 616  
QY 758 RIM 760  
DB 617 TII 619  
RESULT 31  
O66850 PRELIMINARY; PRT; 705 AA.  
AC O66850  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein AQ\_585.  
GN AQ\_585.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]\_TaxID=63363;  
RP SEQUENCE FROM N.A.  
RC STRAIN=VF5;  
RX MEDLINE=98196666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,  
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
RA "The complete genome of the hyperthermophilic bacterium Aquifex  
RT aeolicus";  
RL Nature 392:353-358 (1998).  
DR EMBL; AB006697; AAC06820.1; -;  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSPII/IIPprotein.  
DR Pfam; PF00263; GSPII\_III; 1.  
DR PRINTS; PR00811; BCTERIALGSPD.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 705 AA; 80771 MW; 2AB9870C1EDD61DF CRC64;  
Query Match 8.7%; Score 336; DB 16; Length 705;  
Best Local Similarity 22.4%; Pred. No. 9.5e-11;  
Matches 128; Conservative 110; Mismatches 206; Indels 128; Gaps 21;  
QY 255 QQHDHIVLKHTLFTTLQRSI-----DVADFKTPQVKTLRLAND 297  
DB 193 REYDLIAVLNNLKIISKETLAFVGVQSSINKLSIKQYTPSAKLYDK-DLG 251  
QY 298 TQLIITAGNWE-----LVNKGAAFGYFTFVLPRKQNLSESGVNNAPKFTFG 345  
DB 252 KIMVIDMAENIEKRLDRVLDLIELMSRTTP-----EKEKSKET-----TPR---- 295  
QY 346 RKISLDFQDVERTILQILAKESGVNIVAS---DSVNGKMTLSKDVDPWDQALDLMQAR 402  
DB 296 -----EIETKVFYFKNRDLLEALSRLKENFSGEVILNI-----DKDFNAIIVTS 340  
QY 403 NLD-MRQQGNIVNIAPREDLLAKDKAPL-----QAERKDIALGALYSQ 444  
DB 341 NRSVKSQVTLT-----KDLTESIDKAYLTKLFYRYVISPYLKKIEPMLSEVEVIT- 395  
QY 445 NFOLKYNNVEFRSIRLONAD--TTGNNTLVSGRGSVLIDPATNLTIVD-----TRSV 498  
DB 396 ---LSVNTDEKELISYKNTPTPATFNEGTLKEKAFV--PFNNAILIKDYPRIEKI 450  
QY 499 IEKFRKLIDELDVPAAQVMIARIVEADGFRDLGVKFGATCKKLNKDTSAFGMGVNS 558

DB 451 REKFKELSEKPI---KIKIRAKLIVEVEKSLRELGLISWRTVFSKAYI----- 495  
QY 559 GFGDDDKWGAETKINLIPITAAANSISLVRAISSGALNL---ELASASELSKTKTLANPRV 615  
DB 496 ---PEFQOGETAFRTVTPQPOQSGLLTFTFORNRLNLEFKLLAYEQEGRAKVAESYV 551  
QY 616 LTQNRKAKTESGYEIPFTVTSIANGGSSNTLEKXAVLGLTVTPNITPDQIIMTV--- 672  
DB 552 ITVNGEPAVSSGLEFPFVTEVSLSGGIANVEPKVESIPVLIITPVLPDGNILLSVLA 611  
QY 673 --KTN--KDSQAQCSAGNQTILCISTKNLNTQAMVENGGLTVIGIYEEDNGNTLTIKVPL 728  
DB 612 RRQINSVQEPFV-TQTLTQKIPVSTSRIDVKIPKNGETWIGGAVEKSDSITESGVPK 670  
QY 729 LGDIPVIGNLFTKTKGKTDRELLIFITPRIM 760  
DB 671 LREVFLLGNLFTKTKGKTDRELLIFITPEII 702  
RESULT 32  
Q9PD52 PRELIMINARY; PRT; 775 AA.  
AC Q9PD52  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE General secretory pathway protein D precursor.  
GN XF1527.  
OS Xylella fastidiosa.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OX Xanthomonadaceae; Xylella.  
RN NCBI\_TaxID=2371;  
RP SEQUENCE FROM N.A.  
RC STRAIN=9a5C;  
RX MEDLINE=20165717; PubMed=10910347;  
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,  
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,  
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
RA Fraga J.S., Franca S.C., Franco M.C., Frohne M., Furian L.R.,  
RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.B., Kitajima J.P.,  
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
RA Marck C.F.M., Miraca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
RA Moon D.H., Ngai M.A., Nascimento A.D.T.O., Netto L.E.S.,  
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
RA Quaggio R.B., Roberto F.G., Rodrigues V., de Rosa A.J.M.,  
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
RA da Silva A.C.R., da Silva A.M., da Silva P.R., Silva W.A. Jr.,  
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,  
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
RT "The genome sequence of the plant pathogen Xylella fastidiosa";  
RL Nature 406:151-159 (2000).  
DR EMBL; AE003982; AAP84336.1; -;  
DR InterPro; IPR001982; Bac\_GSPD.  
DR InterPro; IPR004846; GSPII/IIPprotein.  
DR InterPro; IPR004845; GSPII/IIPprotein.  
DR InterPro; IPR005644; NOLW-like.  
DR InterPro; IPR003522; SecYI\_OMP.  
DR InterPro; IPR000897; SRP54.  
DR Pfam; PF00263; GSPII\_III; 1.

DR Pfam; PF03958; GSPII III N; 3.  
DR PRINTS; PRO0811; BCTERIALGSPD.  
DR PRINTS; PRO1337; TYPE3OMGPROT.  
DR PROSITE; PS00300; SRP54; 1  
DR PROSITE; PS00875; T2SP\_D; 1.  
KW Complete proteome.  
SQ SEQUENCE 775 AA; 80749 MW; 0D3504E6960A132F CRC64;  
  
Query Match 8.7%; Score 333.5; DB 16; Length 775;  
Best Local Similarity 20.4%; Pred. No. 1.5e-10;  
Matches 176; Conservative 126; Mismatches 275; Indels 287; Gaps 31;  
  
QY 28 TDIKVSLPNKQIKVVSF-----DKEIVNPTGFTVSSPA- 62  
DB 39 TDITPTQLPDAVGATVPLPDTTETPTALLSDAGGELPVIRGNGKVIQNTVAATPPSM 98  
  
QY 63 -----RIALDPEQTGISMDOQVLEYADPLLSK-----ISAQNSRRA 99  
DB 99 GVAGKGSATNFEGESQAVVKAILGDMQLQNVFIASGVQGTVLTSTPRKPVSAQALSLL 158  
  
QY 100 RVLNLNKPQCYNTEVRGNKWIIFINESDDTVAPAPAPAKAPAKAQOQRTYQVR 159  
DB 159 EMVLGMN-----NARMTYNNGRYSIVQADQALAGTVAPST-----APPAVARG-----FEVR 205  
  
QY 160 SIRIOTLYPGKTTAAAPFTESVSVSAPFSPAKQQAASAKQOOTAAPAKQOATAAPAKQQA 219  
DB 206 VV-----PLKVISASEMKK-----VLDPY 224  
  
QY 220 AAPAKQNIIDFRKDGKQAGIIEALALGAFQOPDISQOHDHIIYTLKNHILPTTLQRLSDV 279  
DB 225 ARENAIVSIDTRN-----LITLAGTR-----VELENYL-----RTVQI 258  
  
QY 280 ADPK-----TPVQKVLTKRLNNDTQLIIITAGNWLWNKSAAPGYFTFOVLPRKQN 330  
DB 259 FVDWLGMMSGVFPIQSGRADKVAADLEKVFGESE-----KTPSAGMERF--MP----- 306  
  
QY 331 LESGGVNAKPTFTGRKISIDFQVEIRTILOILAKESGMNIVASDSVNGKMTL-----SLK 387  
DB 307 LESA---NAVIVITPQASYLD-----QIQKWL-----SVDSVGGALLSFSYALK 348  
  
QY 388 DVPDQALDVMQARNLDMROGNI VNIAPRDELLAKDKAFLOAEKDIALGALYSQNFQ 447  
DB 349 YIKANDLANRLTEVFGVGRREDNSVSLAG-----AQLGVLGSGSG 391  
  
QY 448 LKYNVEEFSIRLNDNADTTGNRTLVSGRG--SVLIDPAT----- 487  
DB 392 DSLPSAGGGSLSAVPSNGGTDNTSSANGGLGGSTLQLSPTQTQNGSVTLFVQGDTVGVS 451  
  
QY 488 -----NTLIVTDRSVIEKPKRLIDELVPAQQVMEARIVEAADGFSRDLGVKFGATGK 542  
DB 452 AVEINTLLVRATPOAWRSIRDVIEKLDVMPQVHIEAQVAEVS----- 495  
  
QY 543 KKLKNDTSAFGWNSGFGGDDKWAETKINLPITAAANSISLVRAISSGA----- 593  
DB 496 --LTNQLS---YGVN-----WFFQNSVNAADAADNGASNGTGICLGAGLPSSAAGRS 542  
  
QY 594 -----LNLELSASELSKTKTLANPRVLTONTKREAKTESGYEI 631  
DB 543 GWKSIAGKVTNGLANWFLGKNAAAINALDQVTOVRLLTQPSVFVRNNAEATLVGARI 602  
  
QY 632 PFTVTSIAN--GGSTNTLKKAVLG--LTVTPNITPDG-----QIMTVKINKDSPAQCA 683  
DB 603 PINSTISNTGLGSNSTSVQVYIDTGVILKVRPRVTKDGMVFLDIVQEVSTPGSLPAACS 662  
  
QY 684 SGNQITL---C---ISTKNLNTQAMVNGGTLIVGGIYEDNGNTLTLYKPLLDIPIVG 736  
DB 663 SASSTLVNSAACNVQINTFRRIKTEAAVSGGDTIMLAGLIDNSGKSGNGVFFLSKVFIVG 722  
  
QY 737 NLFKTRGKKTDRRELLIFITPRIM 760  
DB 723 GLFGTKAQNNNRREIIVLLTPEIV 746

RESULT 33

Q8CVNS PRELIMINARY; PRT; 654 AA.  
ID Q8CVNS5  
AC Q8CVNS5  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Probable general secretion pathway protein D precursor.  
DE YHEF OR C4096.  
GN Escherichia coli O6.  
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
NCBI\_TaxID=217992;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=O6:H1 / CFT073 / ATCC 700528;  
RX MEDLINE=22388234; PubMed=12471157;  
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Rosesch P.,  
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
RT "Extensive mosaic structure revealed by the complete genome sequence  
of uropathogenic Escherichia coli";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
DR EMBL; AE016767; AAN82534.1; -.  
KW Complete proteome.  
SQ SEQUENCE 654 AA; 71147 MW; 2C9822E0B39EDF60 CRC64;  
  
Query Match 8.7%; Score 333; DB 16; Length 654;  
Best Local Similarity 21.5%; Pred. No. 1.3e-10;  
Matches 123; Conservative 97; Mismatches 183; Indels 168; Gaps 17;  
  
QY 351 DFQDVEIRTILOILAKESGMNIVASDSVNGKMTL-----LSLKD----- 388  
DB 35 NFNENADIQFVEIVGQHLGKLTILIDPSVQGTISVRNSDTFSQEQYQFPLSILDLYGSV 94  
  
QY 389 -----VPMQD---ALDLYMQARN-LDM 406  
DB 95 ITLDNGFLKVRSA NVKTS PGMIADSSRPGVGDDELTVRIVLENVAPDLAPLQWMDA 154  
  
QY 407 RQGNINVIAPRDELLAKDKA-----FLQAEKDIALGALYSQNFQKLYKVEEFSIRL 461  
DB 155 GSVGNVHVHPESNVILITGRASTINKLIEVIKRVDPVIGTEKQIHLVEYASAEIDLAEILN 214  
  
QY 462 -----LDNADTTGNRT---LVSG----- 477  
DB 215 QLISESHGSKQMPALLSAKIVADKNTSLIISGPEKARQITSLLSKLDVSEEGNTRV 274  
  
QY 478 -----RG-----SVLIDPATNTLIVTDRSV 498  
DB 275 YVLKYAKATNLVEVLTVGYSEKLDKSGNSRKPSSTSAMDNVAITADEQTNLSLITADQSV 334  
  
QY 499 IEKFRKLIDELVPAQQVMEARIVEAADGFSRDLGVKFG--ATGKKLKNDTSAFGWV 556  
DB 335 QEKLATVIARLDIRRAQVILVEAIIIVEVDQNGNLGVQWANKVGAQQFTN-TGLPVFNA 393  
  
QY 557 NSGFGGDDKWAETKINLPITAAANSISLVRAISSGALN-----LELSASELSKTKTLAN 612  
DB 394 AQGVADYKNGGITSAN-----PANDVFSAYNGVAAGFFNGDWGVLTLASNNKNDILAT 449  
  
QY 613 PRVLTQNRKAEKIESGYEIPFTVTSIANGSGSTNTLKKAVLG--LTVTPNITPDGQIIM 670  
DB 450 PSIVTLDNKLASFNVQDVPVLSGSGSTTSGDVFNTVERKTVGTGTLKVTPOVN-EGDAVL 508  
  
QY 671 TVKINKDSPAQCASGNOTI-LCISTKNLNTQAMVNGGTLIVGGIYEDNGNTLTLYKPL 729  
DB 509 -LEIEQEVSVSSNSTLGPFTFNRTIQNAVIVKTVGTIVVJGLGLDPSKEQSVKVP 567  
  
QY 730 GDIPVIGNLFKTRGKKTDRRELLIFITPRIM 760  
DB 568 GDIPLVQGLFRVTSTTERAKRNLNMFIRPTII 598

## RESULT 35

|                           |       |                   |            |            |
|---------------------------|-------|-------------------|------------|------------|
| Query Match               | 8.5%  | Score 328.5       | DB 16      | Length 690 |
| Best Local Similarity     | 21.8% | Pred. No. 2.4e-10 |            |            |
| Matches 130: Conservative | 116   | Mismatches 228    | Indels 121 | Gaps 19    |

|     |    |  |     |
|-----|----|--|-----|
| 156 | Db | PLIDPRVGVIPTPAAHQVVTOWRSNLERIASLURLDRPSETAGSSSTQVIYLHATA     | 215 |
| 476 | Qy | -----SQRGSVL-IDPATNTLITDTRSVI                                | 499 |
| 216 | Db | SEWVKVLRGLSQGWAPAEVGAEGAXDRPVMAAPGGSGIRLEVEEGTNVVMVGPDSL     | 275 |
| 500 | Qy | EKFKLIDELDVPAQVQWIEARIYAADGSRDLGVKFGATGKKLKNDTSAPGWC--YNS    | 558 |
| 276 | Db | ARYAIRVEQLDIRRAQQVVEAIIAEVSDSSAQELQVQWFADEK-----FGAGIVNF     | 327 |
| 559 | Qy | GFGGDDKWAETKINLPITAAA---NSISLVIRAISS-----GALNLEL--SA         | 600 |
| 328 | Db | GSNG-----VNIANIAGAAAGEDTEALGKLLSATAGATAGIGHGFGGFNAMLVNA      | 378 |
| 601 | Qy | SSLSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSIA--NGGSSTINTELKXAVLGLTV | 658 |
| 379 | Db | LXKSGFNLLSTPTLLTLDNEASILYQGEVPFVTGVTQNNANPYQTTERRKEVGVKLR    | 438 |
| 659 | Qy | TNINTPDGOIMTVKINKDSPAQCASGNQTTILCISTKNLMTQAVWENGTLIVGGIYED   | 718 |
| 439 | Db | KQCIINDNSVRLDIQVEVSSIADSSASDVI--TNKREIKTKVMVEDNGLVILGGLISD   | 496 |
| 719 | Qy | NGNTLTKVPLLGDIPIVGNLFPKTRGKTKDTRRELLIFITPRIM                 | 760 |
| 497 | Db | LSTSNQRVPLLGDIPLYGLRFLFRSDASKNTKQNLVMVFIAPRIL                | 538 |

550 -----SAFGVGVNSGFGDD-----KWGAETKINLPITAAANSISLVRAISSGALNL 596  
 400 PGIVPLAAAGATRSNNGEDDSVLEQARNVAAQSLGL-----SGGLIGLAGQSNDAVFGM 455  
 597 ELSASELSKTKTLANPRVLTONKEAKIESGYIPT-...VTSIANGSGSTNTELKXAV 653  
 456 IIDAVKSDTGSNLLSTPSIMTLDEQARILVQGEVPIITTEVLGAANDNPFRTIQRODVG 515  
 654 LGLTVPTNITPDGQIIMTKINKDSPAQCSAGNOTILCISKNLNTQAMVENGTTLVGG 713  
 516 VELEVRPQINTAGGITLAKEVSAIAGPVSQSESLVFNKQRIETRVVNGAIVLGG 575  
 714 IYEDNGNTLTKVPLGSDIPVIGNLFTKRGKKTDRRELLIPITPRINGTAGNSLR 768  
 576 LLDQNDROTVEKVPILGSDVPGALFRHKSRNRDKNLMLVFIRTIIRDAADAQR 630

RESULT 36  
 Q8F3M6 PRELIMINARY; PRT; 615 AA.  
 ID Q8F3M6;  
 AC Q8F3M6;  
 DT 01-MAR-2003 (TremBLrel. 23, Created)  
 DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)  
 DE General secretory pathway protein D.  
 GN EPSD OR LA2375.  
 OS Leptospira interrogans.  
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
 OX NCBI\_TaxID=173;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;  
 RA Ren S.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE011406; AAN49574.1; .  
 KW Complete proteome.  
 SQ SEQUENCE 615 AA; 68712 MW; 5CD1830F336095A3 CRC64;

Query Match 8.5%; Score 327; DB 16; Length 615;  
 Best Local Similarity 22.5%; Pred. No. 2.5e-10;  
 Matches 122; Conservative 101; Mismatches 192; Indels 128; Gaps 18;

QY 327 KKQNLJESGVNN---APKTFGRKTSIDFQVEIRTLQILAKESGMNIVASDSVNGK-- 381  
 DB 53 KKTSAKTSVTOEPESEKTFVA-----NWRDTLNDLFLKMSAILKKNILLDES LGKKI 107  
 QY 382 MTLSLKDVFWDOA---LDLVMQARNLDRQGNIVNIAPRELLAKAKAF-----LQAEK 433  
 DB 108 TTISQKEPIKNGFIPMKSVLESFGVVEPDLISVVKIDALRSPIVRVGKELIPEE 167  
 QY 434 DIADLGALYQNFOLKYNVEFRSILR---LDNADTTGNRNT---LVSGRGS----- 480  
 DB 168 EVGDVETI--TVIPIENVKPELEPIKRLTSPNTDVIYVRTNTIVLSGSAADINKLLV 226  
 QY 481 -----VLIDPAT----- 487  
 DB 227 LVSEFDVKIEEATPGSISAGDIHIYTYLESAEKIAATLVKLNPNVTQSEDLSGERKPP 286  
 QY 488 -----NTLIIVTDRSVIEKPKLIDELVPAQQVMIEARIVEAAD 527  
 DB 287 PPGQPMKPVDKIKAVGHKESNSVITATNAEWAERIKIIVLDSARKQVLLVILVELTS 346  
 QY 528 GSRDLGVKFGATGKKLKNDSATSGWGVNSGFGGD--DWGAETKINLPITAAANSISL 585  
 DB 347 SOLNDFGIDWRVYKE-----AFG-QFNLSGLSKEANIINSGOVNPNINTLSGFSLGF 397  
 QY 586 VRAISSGALNLSASELSKTKTLANPRVLTONKEAKIESGYEIPFTVTSIANGSGSST 645  
 DB 398 LKAGSEIIGI--LSANQGNENFNVLSPQVLTVDNQEAIEISGVQDVPVYRQSRNAGTGT 456  
 QY 646 NT-----ELKXAVLGLVTPNITPDGQIIMTV--KINKDSPAQCSAGNQTILCISTKNL 697  
 DB 457 NAWTVNDVNEYRPTGKIKFTPHVKNKNKITLFLQEIKNIAEIALAGGNPT---FNRREI 513

QY 698 NTOAMVENGTTLVGGIYVEEDNGNTLTKVPLGSDIPVIGNLFTKRGKKTDRRELLIFITP 757  
 DB 514 KTSISIENTQSVIGGLISNDKQRIIKIIPFGDIPYLGHLFKRTTEKIKKTNLWVFITP 573  
 QY 758 RIM 760  
 DB 574 HIL 576

RESULT 37  
 Q8GBE6 PRELIMINARY; PRT; 658 AA.  
 ID Q8GBE6;  
 AC Q8GBE6;  
 DT 01-MAR-2003 (TremBLrel. 23, Created)  
 DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)  
 DE YtsLD protein.  
 GN YtsLD.  
 OS Yersinia enterocolitica (type 0:8).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Yersinia.  
 OX NCBI\_TaxID=34054;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WA-314;  
 RA Iwobi A.; Rakin A.; Heeseemann J.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ344214; CAC83029.1; .  
 KW SEQUENCE 658 AA; 71592 MW; F91539A6D64230B3 CRC64;

Query Match 8.5%; Score 325.5; DB 2; Length 658;  
 Best Local Similarity 22.6%; Pred. No. 3.3e-10;  
 Matches 166; Conservative 114; Mismatches 284; Indels 169; Gaps 26;

QY 64 IALDPEOTGISMDQOVLEVADPLLSKISAAQNSSRARLVNLNKKPQYNTVEGRGNKWFIF 123  
 DB 7 ITFVYQNRILISQVLSFAVILLVLTALHLPRANAETFSVNFK---NTD----- 54  
 QY 124 INESDDTVS-----APAPPAVKAAPAPAKOQGGCTVYQVRSIRIQTLYPKTTAAAPF 177  
 DB 55 INEFINTVSKNLNKTAIIDPAVKGN-----ISVRS--YOELEPDR--YYPF 96  
 QY 178 TESVSVSAPFSPAKQQAASAKQAQTAAPAKQAQQAAPAKOTNIDFRDKGNA 237  
 DB 97 FLVLEVYG--FTVVM-----PGDVIKIIPAKNTK-----GSAI 129  
 QY 238 GIIEALALGFAGQPDISQQHDHIIIVTLKNHTLP--TTLQSLDVADEFKTPVOKVTLKRLN 295  
 DB 130 PLIE-----GENPAGSDVWVRVSLHNVAAKELAPLLRQLNDAAFGTW-----H 175  
 QY 296 NDTQLIITTAGWELYNKSAAPGYFTFVLPVKQNLSESGVNNAPKFTTGRKISLDFQDV 355  
 DB 176 YDPSNVLTLTGRAAVN-----QLVAIKNVKAGDQ-----TVETIKLOFASA 219  
 QY 356 -ERTILQILAKESGMNIVASDSVNGKMTLSLKDVPWDQALDVMQARNLDRMQGNIVN 414  
 DB 220 SEVARIATESLHSGSKN-----ANGMSAT---IVADERTNSVLIGGEEQVRQ----- 265  
 QY 415 IAPRELLAKDKAPLQAEKDIAADLGALYQNFOLKYNVEFRSILRLNADTT----- 468  
 DB 266 -----MIDTVQELDKQGDIGHG--NTKVIYLVKFAESLDDVLNGVSTNQOGK 311  
 QY 469 GNENTLVSGRGSVLI--DPATNTLIIVTDRSVIEKPKLIDELVPAQQVMIEARIVEAA 526  
 DB 312 GOATPAVMKNNVIVKADACTNSLIINAAPDLLRDLQEVINQLDIRRAQVLEAIIIVEQ 371  
 QY 527 DGRSRLGVKFGATGKKLKNDSATSGWGVNSGFGGDDKMGAEKTNLNPITAAANSISLV 586  
 DB 372 DSDALNLGVQ-----WPNRNGGSGNFPENGASASSITNDMGSLKGI 414

Db 300 -----PSAAGAGGGRNEI-----NIMAHTDTNALV 326  
QY 492 VDTSTRVIEKFKLIDELVPAQVMTIEARTVEADGFSRDLGVKFGATGKKLKNDTSA 551  
Db 327 ISAPDQMRITIESVINQLDITRAQVILVEAIIIVEAGDNGVFGVQWAA----KAGGGTQF 382  
QY 552 FGWGVNSFGGDDKWAETK-----INLP-----ITAAANSILVRAIS-- 590  
Db 383 NNLPTTIGEIGAGIWAQDKEGTITNPSTGEVICQNPKTGKDVITLLAQAALGKNGMAWG 442  
QY 591 -----SGALNLELSASLSLTKTTLANPRVLJTONRKEAKIESGYEIPFTVTSIANGSSST 645  
Db 443 VAMGDFGALVQVQASAD---TNSNVLATPSITTLNQASFIIVGDEVPILTGSTASSNNSN 499  
QY 646 ---NTEKKAVLGLTVPNTIPDQIIMTKINKDSPAQASGNQTI--LCISTKLNLTQA 701  
Db 500 PFQTVRKEVGKLVKVPQIN--EGN---AVKLAIEQEVSGVNGTGVDFISFATRLTIV 555  
QY 702 MVENGGLIIVGIIYEEDNGNTLTQVPLGLDIPVIGNLFKTRGKKTDRRELLIFITPRIM- 760  
Db 556 MADSGIIVVGLLINEEVQESIQVPLGDPILGHLFVSKSSSKTKKQMLMFIKPTIIR 615  
QY 761 -----GTAGNSLRY 769  
Db 616 DGVTMEGIAGRKXNY 630  
RESULT 39  
Q32566 PRELIMINARY; PRT; 585 AA.  
ID Q32566 AC Q32566  
DT 01-JAN-1998 (TRENBLrel. 05, Created)  
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE ETPD protein.  
GN ETPD.  
OS Escherichia coli O157:H7.  
OC Plasmid p0157.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OC NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97237700; PubMed=9084155;  
RA Schmidt H., Henkel B., Karch H.;  
RT "A gene cluster closely related to type II secretion pathway operons  
of Gram-negative bacteria is located on the large plasmid of  
enterohemorrhagic Escherichia coli O157 strains.";  
RL FEMS Microbiol. Lett. 148:265-272(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98290540; PubMed=9628576;  
RA Makino K., Iehii K., Yasunaga T., Hattori M., Yokoyama K.,  
RA Yatsudo H.C., Kubota Y., Yamaichi Y., Iida T., Yamamoto K., Honda T.,  
RA Han C., Ohtsubo A., Kasamatsu M., Hayashi T., Kuhara S.,  
RA Shinagawa H.;  
RT "Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an  
enterohemorrhagic Escherichia coli O157:H7 derived from Sakai  
outbreak.";  
RL DNA Res. 5:1-9(1998).  
DR EMBL; Y09824; CAA70955.1; --  
DR EMBL; AB011549; BAA31759.1; --  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSPII/IIIprotein.  
DR InterPro; IPR004845; GSPIIprotein.  
DR InterPro; IPR005644; NOLW-like.  
DR Pfam; PF00263; GSPII\_III; 1.  
DR Pfam; PF03958; GSPII\_III; N.3.  
DR PRINTS; PR00811; BCTERIALGSPD.  
DR PROSITE; PS00875; T2SP\_D; 1.  
KW Plasmid.  
SQ SEQUENCE 585 AA; 63614 MW; 62AE17CAD87A24FC CRC64;

QY 587 RAISSGALNLELS-----ASELSKTKTLANPRVLJTONRKEAKIESGYEIPFTVTSIANGG 642  
Db 415 TGLATGYRGNSGLTALRTNSQNDILATPSIVTLDNNEAFSGVQEPVL-----SG 468  
QY 643 SSTNT-----ELKAVLGLTVPNTIPDQIIMTKINKDSPAQASGNQTI--CI 692  
Db 469 SQTITGDNIFRTVDRKSVGKLVKVPQINKGDSVLLIEIQEVSSVAERKAPGGTGLGATF 528  
QY 693 STKNLTQAVWENGGLIIVGIIYEEDNGNTLTQVPLGLDIPVIGNLFKTRGKKTDRRELL 752  
Db 529 NTRMVKNVAVGNNIIVVGGGLDSTSHDVTSKVPLGLDIPGIGFLFRSTQRMVKNLUM 588  
QY 753 IFITPRIMTAGN 765  
Db 589 LFIRPTIIEQGD 601  
RESULT 38  
Q8EK9 PRELIMINARY; PRT; 704 AA.  
ID Q8EK9 AC Q8EK9  
DT 01-MAR-2003 (T-EMBLrel. 23, Created)  
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
DE General secretion pathway protein D.  
GN GSPD OR S00166.  
OS Shewanella oneidensis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
OC Alteromonadaceae; Shewanella.  
OX NCBI\_TaxID=70863;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MR-1;  
RX MEDLINE=2297686; PubMed=12368813;  
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
RA Meyer T., Teapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
RA DaRoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
RA Madupu R., Peterson J.D., Unayam L.A., White O., Wolf A.M.,  
RA Vamathevan J., Weidman J., Impraim J., Lee K., Berry K., Lee C.,  
RA Mueller J., Khouri H., Gill J., Ufferback T.R., McDonald L.A.,  
RA Feldblyum T.V., Smith H.O., Venter J.C., Neillson K.H., Fraser C.M.;  
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
Shewanella oneidensis.";  
RL Nat. Biotechnol. 20:1118-1123(2002).  
DR EMBL; AE015466; AAN53253.1; --  
DR TIGR; S00166; --  
KW Complete proteome.  
SQ SEQUENCE 704 AA; 76859 MW; 647B528761A401BF CRC64;  
Query Match 8.4%; Score 323.5; DB 16; Length 704;  
Best Local Similarity 22.9%; Pred. No. 4.8e-10;  
Matches 141; Conservative 104; Mismatches 213; Indels 157; Gaps 27;  
QY 242 LAALGFAGQDIIQQDHIIIVTLNHTLPTTLQSLD-----VADFKTPVQ 287  
Db 86 LQVYGA-----IVEMENNVIKIKDKDAKTAARVANDNDPLGDEMVRIVALYNTEAK 141  
QY 288 KVT--LKRLNN-----DTQLIITAGNWLKNSA-----APGYFTFQVLPKK 328  
Db 142 QLAPLRLQNDNAGGNVNYDFSNVLMLSGRAAVNKLVIIRRYVDKQDTSVQVVP-- 199  
QY 329 QNIESGVNNAKPTFTGRKISLDFQDVEIRTIILQIAKSGM-----NIVASDSVNGKMT 383  
Db 200 --LEYAS-----AGEMVRI-----IDTLRATANQSQLPGQAPKVVADERINA-VV 242  
QY 384 LSLKDYVWDQALDVLVQARNLMRQ--QGN-----IVNIAPRDELLAKDKAF---LQAEKD 434  
Db 243 VSGDESKRQVRELI---HRLDRAEQASTGKTVRYRYAKAEDLVEVLGTFQAKLESEKD 299  
QY 435 IADLGALYSQNFQKLVKNVBEFRESILRLDNDTTG---NRNTLVSGRGSVLIDPATNTI 491



Query Match 8.4%; Score 322.5; DB 2; Length 585;  
Best Local Similarity 22.3%; Pred. No. 4.2e-10;  
Matches 128; Conservative 104; Mismatches 218; Indels 125; Gaps 21;  
QY 242 LAALGAGAGQDISQOHDHIIIVTL-----KNHTLPTTLQRLS-----DVADFPTVQKVT-- 290  
Db 15 LDVVGFA-----VVDMHGILKVVRSKDAKTSAPVVASDVSPGTGDEVTWTRVVPVSVNAAR 70  
QY 291 -----LKRLLNDT---QLIITTAGNHELNVKSAAPGYFTFQVLPKQNLKESGGVNNAPKT 342  
Db 71 DLAPLLRQLNDNAGAGSVVHYEPNSVLLMTGRAAVMKRLMEIVERVDKVGNSRVATVPLT 130  
QY 343 FTGRKISLDFQDVEIRTILOILAKESGMNIV-----ASDSVNGKMTLSLKDVDPWQA 394  
Db 131 YASA-----TDVARLVTELTKETDTAIPAWMTAKLVADERTNSVLVSGEPISSQORI 182  
QY 395 LDLVMOARNLDMRQ--QGNIVNIAPRDELLAKKAFLOAEKDIAIDLALYQNFQKLYKN 452  
Db 183 ISIIKQ---LDROEDVQGN-----TKVIYLYKYAK 208  
QY 453 VEEFRSIL-----RLDNADTTGNRNTLVSGRG-SVLIDPATNTLIIVTDRSVIEKFKLI 506  
Db 209 AKDLVEVLGTGSSSIEN-DSKKSPSTEALRGKVTIKSHEQTNALILTGAQDVIRDLNVI 367  
QY 507 DELDVPAAQVMIEARIVEAADGFSRDLGVKFGATGKKLKNKDTSAFGWGVNSGFGGDDKW 566  
Db 268 SOLDIRRPQVLVEAIIAEIQDADGLNLGIQ-----W-VNKHAG---V 305  
QY 567 GAETKINLPITA-----AANSISLVRAISSGALN-----LELSASESLSKTKT 609  
Db 306 AQFTSTGLPITTMVQTRQNEILDSDQSNALSMFNGIAAGFYQGNWAMLLTALSTSSKNDI 365  
QY 610 LANPRVLTQNRKEAKIESGYEIP-FTVTSIANGSSSTNTLKKAV-LGLTVTNITPDGQ 667  
Db 366 LATPSIVTLDNMEATFNVGQEVPLVSGQTTSGDNI FNTVERKTVGIKLVKPKQINEGDS 425  
QY 668 IIMTVKINKDSPACASGNQITL--CISTKNLNTQAMVNGGTLVGGIYEEEDNGNTLTK 725  
Db 486 VPLLGDIPVLGHLFRKSEQTAQRNLMFLFIRPTII 520  
RESULT 40  
Q9ZGU0 PRELIMINARY; PRT; 642 AA.  
AC Q9ZGU0;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Type II secretion protein.  
GN ETPD.  
OS Escherichia coli O157:H7.  
OG Plasmid p0157.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]\_SEQUENCE FROM N.A.  
RP STRAIN=EDL933;  
RC MEDLINE=98391744; PubMed=9722640;  
RX Burland V., Shao Y., Perna N.T., Plunkett G., Sofia H.J.,  
RA Blattner F.R.;  
RT "The complete DNA sequence and analysis of the large virulence plasmid  
of Escherichia coli O157:H7";  
RL Nucleic Acids Res. 26:4196-4204(1998).  
DR EMBL AF074613; AAC70101.1;  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSPi/II protein.  
DR InterPro; IPR004845; GSPiI protein.  
DR InterPro; IPR005644; NOLW-like.

DR Pfam: PF00263; GSPiI\_III; 1.  
DR Pfam: PF03958; GSPiI\_III\_N; 3.  
DR PRINTS: PR00811; BCTERIAIGSPD.  
DR PROSITE: PS00875; T2SP\_D; 1.  
KW Plasmid.  
SQ SEQUENCE 642 AA; 69911 MW; FBE574CC1DC2B4DC CRC64;  
Query Match 8.4%; Score 322.5; DB 2; Length 642;  
Best Local Similarity 22.3%; Pred. No. 4.8e-10;  
Matches 128; Conservative 104; Mismatches 218; Indels 125; Gaps 21;  
QY 242 LAALGAGAGQDISQOHDHIIIVTL-----KNHTLPTTLQRLS-----DVADFPTVQKVT-- 290  
Db 72 LDVVGFA-----VVDMHGILKVVRSKDAKTSAPVVASDVSPGTGDEVTWTRVVPVSVNAAR 127  
QY 291 -----LKRLLNDT---QLIITTAGNHELNVKSAAPGYFTFQVLPKQNLKESGGVNNAPKT 342  
Db 128 DLAPLLRQLNDNAGAGSVVHYEPNSVLLMTGRAAVMKRLMEIVERVDKVGNSRVATVPLT 187  
QY 343 FTGRKISLDFQDVEIRTILOILAKESGMNIV-----ASDSVNGKMTLSLKDVDPWQA 394  
Db 188 YASA-----TDVARLVTELTKETDTAIPAWMTAKLVADERTNSVLVSGEPISSQORI 239  
QY 395 LDLVMOARNLDMRQ--QGNIVNIAPRDELLAKKAFLOAEKDIAIDLALYQNFQKLYKN 452  
Db 240 ISIIKQ---LDROEDVQGN-----TKVIYLYKYAK 265  
QY 453 VEEFRSIL-----RLDNADTTGNRNTLVSGRG-SVLIDPATNTLIIVTDRSVIEKFKLI 506  
Db 266 AKDLVEVLGTGSSSIEN-DSKKSPSTEALRGKVTIKSHEQTNALILTGAQDVIRDLNVI 324  
QY 507 DELDVPAAQVMIEARIVEAADGFSRDLGVKFGATGKKLKNKDTSAFGWGVNSGFGGDDKW 566  
Db 325 SOLDIRRPQVLVEAIIAEIQDADGLNLGIQ-----W-VNKHAG---V 362  
QY 567 GAETKINLPITA-----AANSISLVRAISSGALN-----LELSASESLSKTKT 609  
Db 363 AQFTSTGLPITTMVQTRQNEILDSDQSNALSMFNGIAAGFYQGNWAMLLTALSTSSKNDI 422  
QY 610 LANPRVLTQNRKEAKIESGYEIP-FTVTSIANGSSSTNTLKKAV-LGLTVTNITPDGQ 667  
Db 423 LATPSIVTLDNMEATFNVGQEVPLVSGQTTSGDNI FNTVERKTVGIKLVKPKQINEGDS 482  
QY 668 IIMTVKINKDSPACASGNQITL--CISTKNLNTQAMVNGGTLVGGIYEEEDNGNTLTK 725  
Db 483 VLEIEQEVSGVADTAVATTDLGATFNTVTNAMLVNGETVTVVGGILLDKSIRGSESK 542  
QY 726 VPLLGDIPVLGHLFRKSKTDRRELLIFITPRIM 760  
Db 543 VPLLGDIPVLGHLFRKSEQTAQRNLMFLFIRPTII 577  
Search completed: December 9, 2003, 10:27:13  
Job time : 49 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 10:20:12 ; Search time 18 Seconds

(without alignments)

2009.085 Million cell updates/sec

Title: US-09-701-271A-2

Perfect score: 3848

Sequence: 1 MNTKLTKIISGLFVATAAFQ.....ELLIFITPRIMTAGNSLRY 769

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID        | Description         |
|------------|--------|-------------|--------|--------------|---------------------|
| 1          | 2762.5 | 71.8        | 711    | 1 OMC_NEIGO  | P35819 neisseria g  |
| 2          | 984.5  | 25.6        | 714    | 1 PILQ_PSAE  | P34750 pseudomonas  |
| 3          | 561.5  | 14.6        | 412    | 1 HOFO_ECOLI | P34749 escherichia  |
| 4          | 516    | 13.4        | 445    | 1 COME_HABIN | P31772 haemophilus  |
| 5          | 439.5  | 11.4        | 430    | 1 VG43_BPBF3 | P03668 bacterioph   |
| 6          | 373.5  | 9.7         | 678    | 1 GSPD_AERHY | P31780 aeromonas h  |
| 7          | 370.5  | 9.6         | 660    | 1 GSPD_KLEBP | P15644 klebsiella   |
| 8          | 364.5  | 9.5         | 678    | 1 GSPD_AERSA | P45778 aeromonas s  |
| 9          | 346.5  | 9.0         | 650    | 1 GSPD_ERWCA | P31701 erwinia car  |
| 10         | 337    | 8.8         | 658    | 1 GSPD_PSAE  | P35818 pseudomonas  |
| 11         | 332    | 8.6         | 650    | 1 GSPD_ECOLI | P45758 escherichia  |
| 12         | 330    | 8.5         | 712    | 1 GSPD_ERWCH | P31700 erwinia chr  |
| 13         | 328    | 8.5         | 710    | 1 GSPD_ERWCH | Q01585 erwinia chr  |
| 14         | 309.5  | 8.0         | 674    | 1 GSPD_VIBCH | P45779 vibrio chol  |
| 15         | 309    | 8.0         | 428    | 1 VG4_BP122  | P15420 bacterioph   |
| 16         | 305.5  | 7.9         | 759    | 1 GSPD_XANCP | P29041 xanthomonas  |
| 17         | 298.5  | 7.8         | 426    | 1 VG4_BP11   | P03666 bacterioph   |
| 18         | 295.5  | 7.7         | 426    | 1 VG4_BP13   | P03665 bacterioph   |
| 19         | 292.5  | 7.6         | 426    | 1 VG4_BP12   | P03664 bacterioph   |
| 20         | 292.5  | 7.6         | 429    | 1 VG4_BP1FL  | OS0300 bacterioph   |
| 21         | 290    | 7.5         | 437    | 1 VG4_BP1FL  | P03667 bacterioph   |
| 22         | 284.5  | 7.4         | 607    | 1 YSCC_YEREN | Q01244 yersinia en  |
| 23         | 283.5  | 7.1         | 701    | 1 HRPP_PSESY | Q01723 pseudomonas  |
| 24         | 268    | 7.0         | 607    | 1 HRPA_XANCV | P80151 xanthomonas  |
| 25         | 240.5  | 6.2         | 568    | 1 HRAI_RALSO | Q52498 ralsstonia s |
| 26         | 227    | 5.9         | 423    | 1 Y4XJ_RHISN | P55702 rhizobium s  |
| 27         | 213.5  | 5.5         | 566    | 1 MXID_SHISO | Q55293 shigella s   |
| 28         | 212.5  | 5.5         | 566    | 1 MXID_SHIFL | Q04641 shigella fl  |
| 29         | 178.5  | 4.8         | 562    | 1 INVG_SALTY | P35672 salmonella   |
| 30         | 147.5  | 3.8         | 2660   | 1 YEEJ_ECO57 | Q8x8v7 escherichia  |
| 31         | 147    | 3.8         | 3421   | 1 TEGU_HSVB  | P28955 equine herp  |
| 32         | 145.5  | 3.8         | 552    | 1 BPPB_ECO11 | Q9s142 escherichia  |
| 33         | 145    | 3.8         | 553    | 1 BPPB_ECO27 | Q47068 escherichia  |

34 144.5 3.8 651 1 E2BD\_YEAST P12754 saccharomyc  
35 144.5 3.8 739 1 PO21\_CHICK P15143 gallus gall  
36 144.5 3.8 1018 1 SCA4\_RICJA Q9aj79 rickettsia  
37 144.5 3.8 1022 1 SCA4\_RICPR Q9jd49 rickettsia  
38 144 3.7 1654 1 OMPB\_RICRI Q53047 r outer mem  
39 142.5 3.7 1065 1 SED4\_YEAST P23365 saccharomyc  
40 142.5 3.7 1185 1 MAPX\_DROME P23226 drosophila  
41 142 3.7 1381 1 YBE7\_YEAST P34216 saccharomyc  
42 141.5 3.7 1011 1 SCA4\_RICAF Q9aj83 rickettsia  
43 141.5 3.7 1139 1 SRE2\_CRIGR Q60429 cricetus  
44 140 3.6 927 1 IF2\_STRAS Q9zf20 streptococ  
45 140 3.6 927 1 IF2\_STRAS Q8ei13 streptococ

#### ALIGNMENTS

RESULT 1  
OMC\_NEIGO  
ID OMC\_NEIGO STANDARD; PRT; 711 AA.  
AC P35819;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE Outer membrane protein OMC precursor.  
GN OMC.  
OS Neisseria gonorrhoeae.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
OC Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=485;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=2686;  
RX MEDLINE=89339707; PubMed=2503445;  
RA Tsai W.M., Larsen S.H., Wilde C.E.;  
RT "Cloning and DNA sequence of the omc gene encoding the outer membrane  
RT protein-macromolecular complex from Neisseria gonorrhoeae.";  
RL Infect. Immun. 57:2653-2659(1989).  
CC -!- SUBUNIT: FORMS A HIGH MACROMOLECULAR COMPLEX IN THE OUTER  
CC MEMBRANE.  
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE, ASSOCIATED TO THE MEMBRANE  
CC THROUGH ITS C-TERMINAL.  
CC -!- SIMILARITY: BELONGS TO THE PULD/OUTD/EXED/XPSD FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL; L19944; AAA25456.1;  
DR InterPro; IPR004846; GSP11/IIIprotein.  
DR InterPro; IPR004845; GSP11proteinc.  
DR InterPro; IPR005644; NOLW-like.  
DR Pfam; PF00263; GSP11\_III; 1.  
DR Pfam; PF03958; GSP11\_III N; 1.  
DR PROSITE; PS00875; T2SP D; FALSE NEG.  
KW Transport; Outer membrane; Signal.  
FT SIGNAL 1 24 POTENTIAL.  
FT CHAIN 25 711 OUTER MEMBRANE PROTEIN OMC.  
SQ SEQUENCE 711 AA; 77483 MW; 249CF9D8DB65F90 CRC64;  
  
Query Match 71.8%; Score 2762.5; DB 1; Length 711;  
Best Local Similarity 75.9%; Pred. No. 58-150;  
Matches 575; Conservative 30; Mismatches 104; Indels 49; Gaps 5;  
  
QY 1 MNTKLTKIISGLFVATAAFQTAAGTASAGNITDKVSSLPNKQIKVKVDFDKIENVNPTGFTVSS 60  
Db 1 MNTKLTKIISGLFVATAAFQTAAGTASAGNITDKVSSLPNKQIKVKVDFDKIENVNPTGFTVSS 60  
QY 61 PARIALDFEQTIGSDMQQVLEVDAPLLSKISAQNSSRARLVNLNPKGQYNTEVRGNKV 120

```
Db 61 PARIALDPEQTGISMDOQVLEYADPLLSKISAQNSSRARLVNLNKPQGYNTEVRGNKV 120
Qy 121 WIFINESDDTVSAPAPPAVPAKAAAPAKQOQCRVTYQVRSIRIOTLYPGKTTAAAFPTES 180
Db 121 WIFINESDDTVSAPAPPAVPAKAA-LRTGKTTCGRTYVRSIRIOTLYPGKTTG----- 172
Qy 181 VVSVSAPFSPAKQOAAASAKQOATAAPAKQOAAAPAKQTNIDPRKQGNAGII 240
Db 173 -----SGIKGNQVRFRRQKCHYR----- 194
Qy 241 ELAALGAGAGOPDI:SQQHDHIIIVTLKNHTLPTTLQSLDVAADFPTKQVTLKRLNNDTQL 300
Db 195 -IGGIGLGGOPDI:SQQHDHIIIVTLKNHTLPTTLQSLDVAADFPTKQVTLKRLNNDTQL 253
Qy 301 IITTAGNWLNVKSAAGPYTFQVLPKPKQNLSEGGVNAKPTTGKISLDFQDVIRTI 360
Db 254 IITTAGNWLNVKSAAGPYTFQVLPKPKQNLSEGGVNAKPTTGKISLDFQDVIRTI 313
Qy 361 LQILAKESGMNIVASDSVNGKMTLSLXDPWDQALDLMQARNLDMRQOQNIYNIAPRDE 420
Db 314 LQILAKESGMNIVASDSVNGKMTLSLXDPWDQALDLMQARNLDMRQOQNIYNIAPR-R 372
Qy 421 LLAQKXAFLOAKDIADLGALYSQNFOLKYKNVEEPRSILRLDNADTTGNRNTLVSGRGS 480
Db 373 AACQORQLLTSGKXHCGRSRAVPFKPLPIETQKGRIPQHPALDNADTTGNRNTLVSGRGS 432
Qy 481 VLIDPATNTLIVTDRSVIEKFKRLIDELDVPAQQVMIEARIVEAADGFSRDLGVKFGAT 540
Db 433 VLIDPATNTLIVTDRSVIEKFKRLIDELDVPAQQVMIEARIVEAADGFSRDLGVKFGAT 492
Qy 541 GKXKLKNDTSAGFVNGSGGGDKWAEKINLPITAAANSISLVRPAISGAINLELSA 600
Db 493 GRXKLKNETSAFGVNGSGGGDKWAEKINLPITAAANSISLVRPAISGAINLELSA 551
Qy 601 SESLSKTKTLANPRVLQNRKEAKIESGYEIPFTVTSIANGSGSSTNTELKAVILGTVTP 660
Db 552 SESLSKTKTLANPRVLQNRKEAKIESGYEIPFTVTSIANGSGSSTNTELKAVILGTVTP 611
Qy 661 NITPDGQIIMTVKINKSPACASGNTILCISTNLNTQAMVNGTGLVGGIYEEDNG 720
Db 612 NITPDGQIIMTVKINKSPACASGNTILCISTNLNTQAMVNGTGLVGGIYEEDNG 671
Qy 721 NTLTKVPLLDGIPVIGNLFKTRGKKTDRRELLIPIPR 758
Db 672 NTLTKVPLLDGIPVIGNLFKTRGKKTDRRELLIPIPR 709

RESULT 2
ID_PSEAE STANDARD; PRT; 714 AA.
AC P34750;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fimbrial assembly protein pilQ precursor.
GN PILO OR PA5040.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25102 / PAK;
RX MEDLINE=94049125; PubMed=7901733;
RA Martin P.R., Hobbs M., Free P.D., Jeske Y., Mattick J.S.;
RT "Characterization of pilQ, a new gene required for the biogenesis of
RL type 4 fimbriae in Pseudomonas aeruginosa";
RN Mol. Microbiol. 9:857-868 (1993).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
```

```
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goffry D., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen";
RL Nature 406:959-964 (2000).
CC -!- FUNCTION: ESSENTIAL FOR THE FORMATION OF PILI. INVOLVED IN THE
CC BIOGENESIS OF TYPE 4 FIMBRIAE PROBABLY BY SERVING AS A "POTHOLE"
CC ALLOWING PASSAGE OF THE FIMBRIAE THROUGH THE OUTER MEMBRANE.
CC -!- SUBCELLULAR LOCATION: Outer membrane (Probable).
CC -!- SIMILARITY: BELONGS TO THE PILQ/OUTD/EXED/XPSD FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L13865; AAA16704.1; ALT_INIT.
CC EMBL: AB004917; AAG08425.1; -.
CC PIR: A83016; A83016.
CC PIR: S37345; S37345.
CC InterPro: IPR001775; Bac GSPD.
CC InterPro: IPR004846; GSP11/IIProtein.
CC InterPro: IPR004845; GSP11/protein.
CC InterPro: IPR005644; NOLW-like.
CC Pfam: PF00263; GSP11_III; 1.
CC Pfam: PF03958; GSP11_III; 1.
CC PRINTS: PR00811; BCTERIALGSPD.
CC PROSITE: PS00875; T2SP_D; 1.
CC Transprot; Protein transport; Outer membrane; Fimbria; Signal;
KW Complete proteome.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 714 FIMBRIAL ASSEMBLY PROTEIN PILO.
FT CONFLICT 391 391 D -> E (IN REF. 1).
FT CONFLICT 411 411 G -> A (IN REF. 1).
FT CONFLICT 547 558 LSAMEKTNGEI -> PVGDGKDRQRY (IN REF. 1).
SQ SEQUENCE 714 AA; 77378 MW; 9A748EE8286FBAB5 CRC64;

Query Match 25.6%; Score 984.5; DB 1; Length 714;
Best Local Similarity 30.6%; Pred. No. 9.2e-49;
Matches 246; Conservative 153; Mismatches 262; Indels 143; Gaps 20;

Qy 1 MNTKLTKIISGLFVATAAFQTA-SAGNIITDIKVSLSLNPKNQKIVKVSFDKEIVNPTGFVTS 59
Db 1 YNSGLSRL--GIALLAAMFAPALLAADLEKLDVAALPGRVVELKLFQFDEPVAAPRGYTIE 58
Qy 60 SPARIALDPEQTGISMDOQVLEYADPLLSKISAQNSSRARLVNLNKPQGYNTEVRGNK 119
Db 59 QPARIALDLPQVONKLTGKNRSLSVGNTRSVTVVEAKDRTLRILNLTALSSYTTVEGNN 118
Qy 120 VMIFINESDDTVSAPAPPAVPAKAAAPAKQOQCRVTYQVRSIRIOTLYPGKTTAAAPFE 179
Db 119 LFVVVGN-----PAGASV--ASAAPV----- 139
Qy 180 SVVSVSAPFSPAKQOAAASAKQOATAAPAKQOATAAPAKQOAAAPAKOT--NIDFRKDGNA 237
Db 140 -----ASPAPASVAPQIKPKPYVPAGAIRNIDFORGEKE 175
Qy 238 GIIEALALGFAGAGPDISQOHDHIIIVTLKNHTLPTTLQSLDVAADFPTKQVTLKRLNND 297
Db 176 GNVVLDLSDPTLSPIQEGGKIRLDFAKTQLPDALRVLDVKDPAATPVQFNASAAQSDR 235
Qy 298 TOLIITTAGNWE-LVYNKSAAGPYTFQVLP-----KKONLESGGVNAKPTTGK 347
Db 236 TSITIEPSGLYDLYVYQT--DNRLTVSIKPMTTTDAERRKKDNF-----AYTGEK 283
Qy 348 ISLDQDQVEIRTILOILAKESGMNIVASDSVNGKMTLSLXDPWDQALDLMQARNLDMR 407
```

```

Db 284 LSLNFODIVRSVLQIAFDTLNVLASDTVOGNTLRLQNVFWDQALDLVTKGLDKR 343
Qy 408 QQQNIYVNIAPRDELLAKAFLOAEKDIALGALYSQNFQLYKXNVBEFRSLRLDNAPT 467
Db 344 KLGNVLLVAPADIAARERQELAEQAQIAELAPRLRELIQVNYAKAADIAKLFQSVTSD- 402
Qy 468 TGNRNLVSRGSLVDPAINTLVITDTSVIEKFKLIDELDVPAQVMIARIVEAAD 527
Db 403 -GGQEGKEGSGITVDDRTNSIIAQPERIDELRIVSQLDIPVRQMIARIVEANV 461
Qy 528 GFSRDLGVKFGATGKKLKNIDTSAFQWGVNSGPGDDKWAETK-----INLPI 576
Db 462 GYDKSLGVRWG-----GAYHKGWMSGYGKDGNIKIDKEDGMNCGPIAGSCTFPT 510
Qy 577 TAAANSISL-----VRAISSG-----ALNLELSASESLSKTKTLANPRVLTONRK 621
Db 511 TGTSKSPSPFVDLGAKDAGTSIGIGITPDNIILDLQLSAMKGTGCEIVSQPKVTSCKE 570
Qy 622 EAKIESGEIIPFTVTSIANGSGSTNTELKAVLGLVTPTNITPDGQIMTVKINKDSPAQ 681
Db 571 TAKILKGEVFPV---QEASSGATSTSFKEAALSLEVTPQITPDNRIIVEVKVTKDAP-- 625
Qy 682 CASGNQTIL----CISTKNLNTQAMVENGTLIVGGIYEEDNGNTLTQVPLLGDIPIVGN 737
Db 626 ---DYQNLNGVPPINNEVNAKILVNDGETIIVGGVFSNEOSKSEKVPFLGELPYLGR 682
Qy 738 LFXTGRGKTDNR-ELLIFITPRIM 760
Db 683 LPR-ROTVDRKNLLEVLTPRIM 705

RESULT 3
HOFO_ECOLI
ID HOFO_ECOLI STANDARD; PRT: 412 AA.
AC P34749;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein transport protein hofQ precursor.
GN HOFO OR HOPO OR B3391.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RN [2]
RP PRELIMINARY SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92105021; PubMed=1309529;
RA Lobner-Olesen A., Boye E., Marinus M.G.;
RT "Identification of the gene (aroK) encoding shikimic acid kinase I of
RT Escherichia coli.";
RL J. Bacteriol. 174:525-529 (1992).
RN [3]
RP IDENTIFICATION.
RX MEDLINE=94049125; PubMed=7901733;
RA Martin P.R., Hobbs M., Free P.D., Jeske Y., Mattick J.S.;
RT "Characterization of pilQ, a new gene required for the biogenesis of
RT type 4 fimbriae in Pseudomonas aeruginosa.";
RL Mol. Microbiol. 9:857-868 (1993).
CC -1- SUBCELLULAR LOCATION: Outer membrane (Probable).
CC -1- SIMILARITY: BELONGS TO THE PULS/OUTD/EXED/XPSD FAMILY.
CC -----
```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/ebis/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL: U18997; AAA58188.1; -;
CC EMBL: AE000414; AAC76416.1; -;
CC PIR: B65134; B65134.
CC EcoGene: EGI2113; hofQ.
CC InterPro: IPR001775; Bac_GSPD.
CC InterPro: IPR004846; GSP1/II/protein.
CC InterPro: IPR004845; GSP1/proteinC.
CC InterPro: IPR005644; NofW-like.
CC InterPro: IPR005322; SecYII-OMPG.
CC Pfam: PF00263; GSP1_III; 1.
CC Pfam: PF03958; GSP1_III_N; 1.
CC PRINTS: PR00811; BCTERIALGSPD.
CC PRINTS: PR01337; TYPE3OMGPROT.
CC PROSITE: PS00875; T2SP.D; 1.
CC Transport; Outer membrane; Signal; Complete proteome.
FT SIGNAL 18
FT CHAIN 19..412
FT SEQUENCE 412 AA; 44716 MW; 99FEABE2606682ES CRC64;
SQ
Query Match 14.6%; Score 561.5; DB 1; Length 412;
Best Local Similarity 32.7%; Pred. No. 5.1e-25;
Matches 141; Conservative 89; Mismatches 146; Indels 55; Gaps 10;
Qy 346 RKISLDFODVEIRTLQILAKESGNIVASDSVNGKMTLSKDVFWQALDLVQARNLD 405
Db 22 QKVTLMVDVDFVAQVLOALAEQKLNLVSPDVSGTVSLHLDVFWKALQTVVKSAGLI 81
Qy 406 MRQGNIV-----NIAPRELLAKDKAFLOAEKDIALGALYSQNFQLYKXNVKEE 455
Db 82 TRQGNILSVHSIAWQNNIARQEAQRAQANIPLE-----NRSITLQVADAGE 131
Qy 456 FRSILRLDNADTGNRNLTVSGRSLVIDPATNLTIVTDRSVIEKPRKLIIDELDVPAQ 515
Db 132 LAK-----AGEK--LLSAGKSMVDRKTNLLRDNKLTALSLEQWVAQMDLPVQ 180
Qy 516 VMIEARIVEAADGFSRDLGVKFGATGKKLKNIDTSAFQWGVNSGPGDDKWAETKINLP 575
Db 181 VELSAHVITINEKSLRELGVKM-----TLADAQHAGVGQVTLTGSN-----LS 224
Qy 576 ITAAANSISL-VRAISSGALNLELSASESLSKTKTLANPRVLTONRKEAKIESGYEIPFT 634
Db 225 VATATTVGPNIGRNGRLLELSALEQKQQLDIASPRLLASHLQPASIKQSGSEIPYQ 284
Qy 635 VTSIANGSGSTNTELKAVLGLVTPTNITPDGQIMTVKINKDSPAQ-----ASGNQTILC 691
Db 285 VSSGESG--ATSVFKEAVLGMVETPTVLQKGRIRLKLHISQNVPGVQVLOQADGE--VLA 340
Qy 692 ISTKNLNTQAMVENGTLIVGGIYEEDNGNTLTQVPLLGDIPIVGNLTKRGKTDREEL 751
Db 341 IDKQEIETQVEVKSGETLALGGITRKNKSQDSVPLLGDIPIVGNLTKRGKTDREEL 400
Qy 752 LIFTPTIMGT 762
Db 401 VVFTPLRVSS 411

RESULT 4
COME_HAEIN
ID COME_HAEIN STANDARD; PRT: 445 AA.
AC P31772;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Competence protein E precursor (DNA transformation protein comE).
```

25 PKT-DNERFFIRLSQAPLAQTLEQAFQDQVNLVGDILLENKISLKNINDMPRLQIIA 83  
 400 QARNLDMRQQ-----GNIVNIARDELLAKDKAFLOAEKDIALGALY 442  
 84 KSKHJTLNKGDIYVINGSQSGQVAGNLTNEPH-----LV 121  
 443 SQNFQJKNVBEFRSILRLDNADTTGNRTNLVSGRGSVLIDPAINTLIVTTRSVIEKF 502  
 122 SHTVKLHFKAQSELMKSL-----ITGS--GSLSPAGSITTFDRSLLVLIQDEPRSVONI 174  
 503 RKLIDELVPAQVNVTEARIVEAAGFSRDLGVKFGATGKKLKNDTSAFGMGVNSGFGG 562  
 175 KKLIAEMDKPIEQIAEARIIVITDESLEKELGVRCIF--NTENARRVAGSLTNSF-- 230  
 563 DDKWAETKINLPITAA-ANSISL-VRAISSGALNLELSASELSKTKTLANPRVLTONR 620  
 231 -ENIADNLNVNPAITTPAGSIALQVAKINGRLLDLELSALERENNVEIIAGPRLLTTNK 289  
 621 KEAKIESGVEIEPTVTSIANGSSNTTELKKAVLGJVTTPNTIPDQGIIMTVKINKDSP- 679  
 290 KSAISKQGEIPIYVSNTRN--DTQSVREFAVLGLEVTPHISKDNNILDLILVSNQSPG 347  
 348 SRVAYGQNEWSIDKQEIINTQVFAKDGETILVGGVPHDTITKSEDKVPLLGDIPVTKRLF 407  
 740 KTRGKKTDRRELLIFITPRIMGAGNSL 767  
 408 KXESERHKRELVIFVPHIL-KAGETL 434  
 RESULT 5  
 ID43 BPPF3 STANDARD; PRT; 430 AA.  
 AC P03658;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 01-MAR-1989 (Rel. 10, Last annotation update)  
 DE 46.4 kDa protein (ORF 430).  
 DE Bacteriophage Pf3.  
 OS Viruses; ssDNA viruses; Inoviridae; Inovirus.  
 ON NCBI\_TaxID=10872;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=New-York, and Nijmegen;  
 RX MEDLINE=8593231; PubMed=3928901;  
 RA Luitert R.G.M., Putterman D.G., Schoenmakers J.G.G., Konings R.N.H.,  
 RA Day L.A.;  
 RT "Nucleotide sequence of the genome of Pf3, an Incp-1 plasmid-specific  
 RT filamentous bacteriophage of Pseudomonas aeruginosa";  
 RL J. Virol. 56:268-276(1985).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR ENBL; M11912; AAA88381.1; -;  
 DR EMBL; M19377; AAA88390.1; -;  
 DR PIR; A04270; Z4BPJ3.  
 DR InterPro; IPRO01775; Bac\_GSPD.  
 DR InterPro; IPRO04846; GSPiI/IIprotein.  
 DR InterPro; IPRO05644; NOLW-like.  
 DR Pfam; PF00263; GSPiI\_III; 1.  
 DR Pfam; PF03958; GSPiI\_III\_N; 1.  
 DR PRINTS; PR00811; BCTERIATGSPD.  
 DR SEQUENCE 430 AA; 46490 MW; E7D3A7A09AD77C53 CRC64;  
 Query Match 11.4%; Score 439.5; DB 1; Length 430;  
 Best Local Similarity 27.5%; Pred. No. 4.7e-18;

|                       |       |                    |       |             |
|-----------------------|-------|--------------------|-------|-------------|
| Query Match           | 11.4% | Score 439.5;       | DB 1; | Length 430; |
| Best Local Similarity | 27.5% | Pred. No. 4.7e-18; |       |             |

Matches 118; Conservative 93; Mismatches 175; Indels 43; Gaps 10;

QY 340 PKTGTGKISLDFDQVIRTILOTLAKESGNIVASDVNGKMTLSLKDVDPWQALDVLV 399  
 DB 16 PFASFASRLTVKHEIDIRVAIPJVAFCGRSVVLGFSIQGVSLDDDDVPCQAFDILL 75  
 QY 400 QARNLDMRQGNIVNIAPRDELLAKDKA---FLQAEKDADIADLGALYSQNFQKLYKXNVEEF 456  
 DB 76 ESNHLLSSMVGDLVITAMQVNLSEKADDLRTFRDLFNANDIERRVINIVHASASEV 135  
 QY 457 RSLR--LDNADTTGNRTLVSGRGSVLIDPATNTLVDTSTRVIEKRLIDELVPAQ 514  
 DB 136 VSLFKESFMSLDAPGM-----SMTVDERTNSFAALPSSFFPALESVIAIDVPEV 186  
 QY 515 QVMTEARIVAAADGFSRDLGVKFGATKKLKNDTSAFGWGVNSFGGDDKWAETKINL 574  
 DB 187 QVAIEANVVEASVDXRELGLNWG--GALSIGN-----WSAVTAGDL 226  
 QY 575 PITAAANSISLVRAISSALNLE--LSASELSKTKTLANPRVLTQNRKEAKIESGVEIP 632  
 DB 227 SV-AAGSSIGF--GFLSNTLSLGLFTAMENEGNRVRSRPTLLTLDROASVLRGTPEL 283  
 QY 633 FTVT'SIANGSSSTWELKAVLGLTVPNTTPDQGIIMTVKINKDSPAQCASGNQITLCI 692  
 DB 284 YQQA---GDGATSVAFKHAALSLEVXVPISPDNSIVIEVLVSDSP-NFSNAIDGVPI 339  
 QY 693 STKNLTQAMVNGGTLVIGGIYEDNGNTLTKVPLLGDIPIVGNLFKTRGKKTDRRELL 752  
 DB 340 DTRNLVTIRVPHGQTVLGGVYSTINQQGSSRVSGISRIPIGIRLFKKEHVEHQVELL 399  
 QY 753 IFITPRIMG 761  
 DB 400 IFLTPIRLG 408

RESULT 6  
 GSPD\_AERHY STANDARD; PRT; 678 AA.  
 ID GSPD\_AERHY STANDARD; PRT; 678 AA.  
 AC P31780;  
 DT 01-JUL-1993 (Rel. 25, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE General secretion pathway protein D precursor.  
 GN EXPD.  
 OS Aeromonas hydrophila.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Aeromonadales;  
 OC Aeromonadaceae; Aeromonas.  
 OX NCBI\_TaxID=644;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ah65;  
 RA Howard S.P.;  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 408-678 FROM N.A.  
 RC STRAIN=Ah65;  
 RX MEDLINE=92349963; PubMed=1640836;  
 RA Jiang B., Howard S.P.;  
 RT "The Aeromonas hydrophila exeE gene, required both for protein  
 secretion and normal outer membrane biogenesis, is a member of a  
 general secretion pathway";  
 RL Mol. Microbiol. 6:1351-1361(1992).  
 CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE  
 CC EXPORT OF PROTEINS.  
 CC -!- SUBCELLULAR LOCATION: Outer membrane (Probable).  
 CC -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC EMBL: X66504; CAA47124.1; -  
 DR PIR: S22668; S22668.  
 DR InterPro: IPR001775; Bac\_GSPD.  
 DR InterPro: IPR004846; GSPFII/IIprotein.  
 DR InterPro: IPR004845; GSPFIIproteinC.  
 DR InterPro: IPR005644; Noliw-like.  
 DR Pfam: PF00263; GSPFII\_III; 1.  
 DR Pfam: PF03958; GSPFII\_III\_N; 3.  
 DR PRINTS: PRO0811; SCTERIALGSPD.  
 DR PROSITE: PS00875; T2SP\_D; 1.  
 KW Transport; Outer membrane; Signal.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 678 GENERAL SECRETION PATHWAY PROTEIN D.  
 SQ SEQUENCE 678 AA; 72451 MW; 43B33A28861B0238 CRC64;

Query Match 9.7%; Score 373.5; DB 1; Length 678;  
 Best Local Similarity 23.9%; Pred. No. 5.1e-14;  
 Matches 115; Conservative 97; Mismatches 189; Indels 80; Gaps 13;

QY 353 QQVEIRTIQIILAK-----ESGMNIVASDSVN-----GKMTLSLK-----DVPWDQAL 395  
 DB 133 RNVSVRELAPLRLQNDNAGGNNVHYDPSNVLLITGRAAVNRLVEVVRVDKAGDQEV 192  
 QY 396 DLV-----MQARNLDMRQGNIVNIAPRDELLAKDKA----- 427  
 DB 193 DIILKYASAGEVRLVTNLNKGNSQGGNTSLLLAPKVADERTNSVVSVEPKARARI 252  
 QY 428 ---FLQAEKDADIADLGALYSQNFQKLYKXNVEEFSILR-----LDNADTTGNRN 472  
 DB 253 IQMVRDLRDLQSQ--NTRVFLKYKAKAMVEVLKGVSSIEADKKGGGTATTAGGA 310  
 QY 473 TLVSGRGSVLIDPATNTLVDTSTRVIEKRLIDELVPAQQVMIEARIVEAADGFSRD 532  
 DB 311 SIGGGKLAISADETTNALVITRAQPDVMAELQVAKLDIRRAQVLVEAIIIVEIADGDGLN 370  
 QY 533 LGVKFGAT--GKKLKNDTSAFGWGVNSFGGDDKWAETKINLPITA---AANSISLV 586  
 DB 371 LGVQWANTNGGTQFTN-----AGPGISVAIAADYKONGTTTGLAKLAENFNGWA 422  
 QY 587 RAISSGALNLELSASELSKTKTLANPRVLTQNRKEAKIESGYTIIPF-TVTSIANGSST 645  
 DB 423 AGFYQGNWAMLVLTALSTNKSILSTPSIVTMDNKEASPNVQGEVVPVQTGNTSGDIT 482  
 QY 646 NTELKXAVLG--LTVTPNITPDGQIIMTVKINKDSPAQCASGNQTI-LCISTKNLTQAM 702  
 DB 483 FSTIERKTVGTLVVTTPQINEGDSVLLTIEQEVSSVVGQATGTDGLGPTFTRTVKNAYL 542  
 QY 703 VENGGLTIYGGIYEDNGNTLTKVPLLGDIPIVGNLFKTRGKKTDRRELLIIFITPRIMG 762  
 DB 543 VKSGETVVLGGLMDEQTEKESKVPILGDIPIVGLYLFIRSTSNNTSKENLWVIRPTILRD 602  
 QY 763 A 763  
 DB 603 A 603

RESULT 7  
 GSPD\_KLEPN STANDARD; PRT; 660 AA.  
 ID GSPD\_KLEPN STANDARD; PRT; 660 AA.  
 AC P15644;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE General secretion pathway protein D precursor (Pullulanase secretion  
 DE envelope puld).  
 GN PULD.  
 OS Klebsiella pneumoniae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Klebsiella.  
 OX NCBI\_TaxID=573;  
 RN [1]





```
QY 396 DLV-----MOARNLMDRQGNINVIAPRDELLAKOKA----- 427
D 193 DIILRYASAGWVRLTNLNDKQNTQGGNTSLLLAPKVVADERTNSVVSGBPKARARI 252
QY 428 ---FLOAKEDTADLALYSQNFOLKYN---VEEPSILRLDNADTGNRENTLVSGRGS 480
D 253 IQMVRQDRDRLOSQ---NTRFYLKYGKAKDMVEVLKGVSTSPADKKGGTTAGGNAS 310
QY 481 -----VLIDPATNLIVTTRSVIEKFRKLDELDPVQAQVMIEARIVEAADGFSRDL 533
D 311 IGGGLAISADETTNALVITAQPDVMAELQVAKLDIRRAQVLVEAIIVEIADGDGLNL 370
QY 534 GVKEGATCKKKLKNDSAFSGVNGSGGGDDKWAETKINLPI-----T 577
D 371 GVQWANTN-----GGGTQF---TDTNLPICSVIAIAKADYENGT 408
QY 578 AAANSISLVRAISSG-----ALNLELSASESLSKTKTLNPRVLTONRKEAKIESGYE 630
D 409 GLADLAKGFNGMAAGFYHGNWAALVTALSTS---TKSDILSTPSIVTMDNKEASFNVQGE 465
QY 631 IPFTVTSIANGSSNTT-----ELKKAVLGLTVPNTIPDQGIIMTVKINKDSPAQ- 681
D 466 VP-----VQSGSQSTSDQVFNTERKTGTGKLTVPQINEGVSVLNIEQVSSVAQK 520
QY 682 CASGNQTI-LCISTKNLNTQAMVNGGTLIVGGIYEEDNGNTLTKVPLLDGIPVIGNLFX 740
D 521 QATGTADLGPFTDRTIKNAVLKSGETVVLGGLMDEQTBKSKVPLLDGIPVLGYLFR 580
QY 741 TRGKKTDERELLIPITPRIMGTA 763
D 581 STNNTSKRNLMVFIPTILRDA 603

RESULT 9
ID -GSPD_ERWCA STANDARD; PRT; 650 AA.
AC F3A701;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE General secretion pathway protein D precursor (Pectic enzymes
DE secretion protein outD).
GN OUTD.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCRI193;
RX MEDLINE=93316842; PubMed=8326859;
RA Reeves P.J., Whitcombe D., Wharam S., Gibson M., Allison G., Bunce N.,
RA Barallon R., Douglas P., Mulholland V., Stevens S., Walker S.,
RA Salmund G.P.C.;
RT "Molecular cloning and characterization of 13 out genes from Erwinia
RT carotovora subspecies carotovora: Genes encoding members of a general
RT secretion pathway (GSP) widespread in Gram-negative bacteria.",
RL Mol. Microbiol. 8:443-456(1993).
RN [2]
RP IDENTIFICATION OF PROBABLE FRAMESHIFT.
RA Bairoch A.;
RL Unpublished observations (FEB-1997).
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLOCATION OF THE
CC MULTIPLE PECTIC ENZYMES.
CC -!- SUBCELLULAR LOCATION: Outer membrane (Probable).
CC -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL: X70049; CAA49645.1; -
DR PIR: S32858; S32858.
DR InterPro: IPR001775; Bac_GSPD.
DR InterPro: IPR004846; GSPII/IIIprotein.
DR InterPro: IPR004845; GSPIIproteinC.
DR InterPro: IPR005644; NOLW-like.
DR InterPro: IPR003522; SecIII_OMP.
DR Pfam: PF00263; GSPII_III; 1.
DR Pfam: PF03958; GSPII_III_N; 3.
DR PRINTS: PR00811; BCTERIALGSPD.
DR PRINTS: PR01337; TYPE3OMGPROT.
DR PROSITE: PS00875; T2SP_D; 1.
KW Transposite; Outer membrane; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 650 GENERAL SECRETION PATHWAY PROTEIN D.
FT CONFLICT 139 164 ELNDNMRGTCGDYSPANVVMTGRA -> VQRQVAVDW
FT ELNDNMRGTCGDYSPANVVMTGRA (IN REF. 1).
FT RLTCTERRRDDWPR (IN REF. 1).
SQ SEQUENCE 650 AA; 70144 MW; 9A228C369B0E2AFC CRC64;

Query Match 9.0%; Score 346.5; DB 1; Length 650;
Best Local Similarity 23.4%; Pred. NO. 1.7e-12;
Matches 117; Conservative 84; Mismatches 189; Indels 109; Gaps 13;

QY 284 TPVKYKTLRLNNDTOLIITAGNWELV-----NKSAAPGYFTFQVLPKKQNLES 333
D 172 TIVERVDQTDGRNVTPISYASSIEVVKWNLNKNMDEKSALPGMLTANVAVADERTNSA 231
QY 334 GGVNNAKPTFTGRKISLDFQDVEIRTILOILAKESGMNIVASDSVNGKMTLSLKDVPMQ 393
D 232 AG-----FQEPNSR-----QR 242
QY 394 ALDVMQARNLMDRQ---QGNIVNIAPRDELLAKDFAKLAQEKDIADLGALYSQNFOLK 451
D 243 VIDMRQ---LDRQAVQGN-----TKVIYLKVA 268
QY 452 NVBEFRSILR-LDNADTTGNRNTLVSGRGSVLI--DPATNTLIVTDRSVIEKFRKLIDE 508
D 269 KAADLVEVLTVGDSIQTDQNALPALRKDLSIKAHEQTNLSLIVNAAPDIRDLEQVIAQ 328
QY 509 LDVPAQVMIEARIVEAADGFSRDLGVKFG--ATGKKLKNDSATGAGVNSGFGGDDKW 566
D 329 LDIRRFQVIVEAIIAEVQADGNLGVQWANKNAGVTQTNT----GLFITMMAGADQF 384
QY 567 GAETKINLPITAAANSISLVRA-ISSGALNLELSASESLSKTKTLNPRVLTONRKEAKI 625
D 385 RRDGTLGTAATTALGGFNGIAAGFYQGNWGLMTALSSNSKNDILATPSIVTLDNMEATF 444
QY 626 ESGYEIPFTVTSIANGSGSNTTELKXAVLG--LTVTPNTIPDQGIIMTVKINKDSPAQCA 683
D 445 NVGQEVPEVLASGQTTSGDNVFTVERKTVGIKLVKVPQINEGDSVLEIEQEVSVDAA 504
QY 684 SGNQTI--CISTKNLNTQAMVNGGTLIVGGIYEEDNGNTLTKVPLLDGIPVIGNLFT 741
D 505 SSSTNLGATFTRTYNNNAVLYSSGDTVVVGGLLDKSTNESANKVPLLDGIPVLGYLFRS 564
QY 742 RGKKTDERELLIPITPRIM 760
D 565 NSTETKRNLMVFIPTILRDA 583

RESULT 10
GSPD_PSEAF
ID -GSPD_PSEAF STANDARD; PRT; 658 AA.
AC P35818; Q9HZB2;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE General secretion pathway protein D precursor.
DE XCFQ OR PA3105.
```

```
OS Pseudomonas aeruginosa
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=95020542; PubMed=7934833;
RA Akrim M., Bally M., Ball G., Tommassen J., Teerink H., Filloux A.,
RA Lazdunski A.;
RA "Xcp-mediated protein secretion in Pseudomonas aeruginosa:
RT identification of two additional genes and evidence for regulation of
RT xcp gene expression.";
RL Mol. Microbiol. 10:431-443 (1993).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Huynh W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen.";
RL Nature 406:959-964 (2000).
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC EXPORT OF PROTEINS.
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC
CC EMBL; X68594; CAA48582.1; -
CC PIR; AE004734; AAG06493.1; -
CC DR InterPro; IPR001775; Bac GSPD.
CC DR InterPro; IPR004846; GSP11/III protein.
CC DR InterPro; IPR004845; GSP11 protein.
CC DR InterPro; IPR005644; NOLW-like.
CC DR Pfam; PF00263; GSP11_III; 1.
CC DR Pfam; PF03958; GSP11_III_N; 3.
CC DR PRINTS; PR00811; BCTRIALGSPD.
CC DR PROSITE; PS00875; T2SP D; 1.
CC KW Transport; Outer membrane; Signal; Complete proteome.
CC FT SIGNAL 1 34 POTENTIAL.
CC FT CHAIN 35 658 GENERAL SECRETION PATHWAY PROTEIN D.
CC SQ SEQUENCE 658 AA; 69953 MW; EC2F81FD1A185D50 CRC64;
Query Match 8.8%; Score 337; DB 1; Length 658;
Best Local Similarity 23.6%; Pred. No. 5.8e-12;
Matches 147; Conservative 117; Mismatches 210; Indels 148; Gaps 28;
QY 235 KNAGIIELALGAFAGOPDISQDHDIIVLTKNHTLPTTLQSLDVADEKTPQVKVTLKRL 294
Db 35 ENSG-----GNAFVPAGNQEAHTNLKD-----ADREFIDQISEITGTF 77
QY 295 NNDTQIITAGNWLNVK---SAAPGYTF-----QVLPKQ-NLE 332
Db 78 VVDPRV-----KGQSVVSKAQLSLSEVYQLFLSVMTGHTFTVVAQGDQARIVPNAEAKTE 133
QY 333 SGGVNNAKPTFTGRKISLDPQV-EIRTLQILAKESGMNIVASVNGKMTLSLKDVFP 391
Db 134 AGGQAPARLETRVIVQVQSPVSELIPLIRLVPQYG-HLAAPVPSANA-LIISRSANI 191
QY 392 DQALDLVMQARNLDMRQQG-----NIVN-----IAPDELL-----AKQXAFLOAEKD 434
```

DR InterPro: IPR004845; GSP1proteinc.  
DR InterPro: IPR005644; Noliw-like.  
DR InterPro: IPR003522; SecIII-OMPG.  
DR Pfam: PF00263; GSP11\_III; 1.  
DR Pfam: PF03958; GSP11\_III N; 3.  
DR PRINTS: PR00811; BCTERIALGSPD.  
DR PRINTS: PR01337; TYPE3OMGPROT.  
DR PROSITE: PS00875; T2SP\_D; 1.  
KW Transport; Outer membrane; Signal; Complete proteome.  
FT SIGNAL 1 23 POTENTIAL.  
FT CHAIN 24 650 PROBABLE GENERAL SECRETION PATHWAY  
FT PROTEIN D.  
SQ SEQUENCE 650 AA; 70698 MW; 973259A12A7237B2 CRC64;

Query Match 8.6%; Score 332; DB 1; Length 650;  
Best Local Similarity 21.5%; Pred. No. 1.1e-11;  
Matches 123; Conservative 97; Mismatches 183; Indels 168; Gaps 17;

QY 351 DFQDVEITLQILAKESGMNIVASDSVGMGT-----LSLKD----- 388  
DB 31 MNNADIRQFEIVGQHLGKTLIDPSVQGTISVRSNDTFSQGEYQFFLSLIDLYGYSV 90  
QY 389 -----LVPMQD--ALDLVMQARN-LDM 406  
DB 91 ITLNGFLKVRNSNVTSFGMIADSRPGVDELVTRIVPLENVNVPARDAPLLRQWMDA 150  
QY 407 RQQGNIVNIAPREDLLAKDKA-----FLQAEKDIADLGALYSQNFQKYKNVEEFSILR 461  
DB 151 GSVGNVHYBPSNVLITGRASTINKLIEVKRVDVIGTEKQIIHLEYASAEDLAEILN 210  
QY 462 -----LDNADITGNRT-----LVSG----- 477  
DB 211 QLISESHGKSQMPALLSAKIVADKRTNSLIISGPEKARQRTSLLSLSDVEESEGNTRV 270  
QY 478 -----RG-----SVLIDPATNTLIIVTDRSV 498  
DB 271 YLYKAYKATNLVEVLTVGSEKLDKDEKNAKRPSSSGAMDNVAITADQTNLSVITADQSV 330  
QY 499 IEKPKLIDELVPAQVMIEARIVEADGFSRDLGVKFG--ATGKKLKNDSAFGNGV 556  
DB 331 QEKLATIARLDIRRAQVLEAIIVEYQDGNLGLVQWANKNVGAQOFTN-TGLPIFNA 389  
QY 557 NSGFGGDDKGAETKINLPITAAANSISLVRAISSGALN-----LELSASSELSTKTKTLAN 612  
DB 390 AQGVADYKNGGITSAN-----PANDMFSAYNGMAAGFFNGDWGVLTLALASNKNKDILAT 445  
QY 613 PRVLTQNRKEAKIESGVEIPTVTISIANGSSTNTELKAVLG--LTVTNPITPDGQIIM 670  
DB 446 PSIVTLDNKLASFVNGDVPVLGSGQSTTSGDNVNTVERKTVGKLVTPQVN-EGDAVL 504  
QY 671 TVKINKDSPACASGNQTI-LCISTKNLTQAMVENGTLIVGIEEDNGNTLTKVPLL 729  
DB 505 -LEIEQEVSSVDSSNSTLGTFTNTRIQNAVLKVTGTETVVLGSLDDFSKEQSVKPLL 563  
QY 730 GDIPVIGNLPKTRGKTKDRRELLIFITPRIM 760  
DB 564 GDIPVLGQLFRYSTRAKRNLNMFIRPTII 594

RESULT 12  
GSPD\_ERWCH  
ID GSPD\_ERWCH STANDARD; PRT; 712 AA.  
AC P31700;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE General secretion pathway protein D precursor (Pectic enzymes secretion protein outD).  
GN OUTD.  
OS Erwinia chrysanthemi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Pectobacterium.  
OX NCBI\_TaxID=556;

[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-EC16;  
RA MEDLINE=93054355; PubMed=1429461;  
RX Lindeberg M., Collmer A.;  
RT "Analysis of eight out genes in a cluster required for pectic enzyme secretion by Erwinia chrysanthemi: sequence comparison with secretion genes from other Gram-negative bacteria";  
RL Bacteriol 174:7385-7397(1992).  
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLLOCATION OF THE MULTIPLE PECTIC ENZYMES.  
CC -!- SUBCELLULAR LOCATION: Outer membrane (Probable).  
CC -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; LO2214; AAA24831.1;  
DR InterPro: IPR001775; Bac\_GSPD.  
DR InterPro: IPR004846; GSP11/IIIproteinc.  
DR InterPro: IPR004845; GSP1proteinc.  
DR InterPro: IPR005644; Noliw-like.  
DR Pfam: PF00263; GSP11\_III; 1.  
DR Pfam: PF03958; GSP11\_III N; 3.  
DR PRINTS: PR00811; BCTERIALGSPD.  
DR PROSITE: PS00875; T2SP\_D; 1.  
KW Transport; Outer membrane; Signal.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 712 GENERAL SECRETION PATHWAY PROTEIN D.  
FT DOMAIN 288 353 GLY/SER-RICH.  
FT SEQUENCE 712 AA; 76478 MW; 8A065D9ADAE24888 CRC64;

Query Match 8.6%; Score 330; DB 1; Length 712;  
Best Local Similarity 27.9%; Pred. No. 1.6e-11;  
Matches 86; Conservative 80; Mismatches 126; Indels 16; Gaps 8;

QY 464 NADTITNRTLVSGRG-SVLIDPATNTLIIVTDRSVIEKFKLIDELVPAQVMIEARI 522  
DB 347 SANSSGRTIVTQKEVTVRAHDQTNLSLIITPPDIMRDLQVINCLDIRRQVLVEALI 406  
QY 523 VRAADGFSRDLGVKFGATGKKLKNDSAFGNGVSGFGDDKGAETKINLPITAAANS 582  
DB 407 AEIQDADGNLGIQW--ANKRAGMTQFTNTGIPITAMIGTDQFRSDGTLT--TAYASA 461  
QY 583 ISLVRAISSGAL-----NLELSASELSKTKTLANPRVLTQNRKEAKIESGVEIPF---TV 635  
DB 462 LSNFGITAGFRGNWSMLLTALSSDGKNDVLATPSIVTLDNWEATFNVGQEVPLTGSQ 521  
QY 636 TSIANGSSTNTELKAV-LGIVTVPNTPDGQIIMTVKINKDSPAQ-CASGNQTI-LCI 592  
DB 522 TTVGSGDNIFNTVERKTVGKLVKVPQINEGDSVLLQIEQEVSSVAEGSGSSSLGVTF 581  
QY 693 STKNLTQAMVENGTLIVGIEEDNGNTLTKVPLLGDIPVIGNLPKTRGKTKDRRELL 752  
DB 582 NRTVNNAVNTNREIVVVGGLDKTAIETNNKVPLLGDIPWLGSLFRSKTQTMSENLM 641  
QY 753 IFITPRIM 760  
DB 642 LFLRPTII 649

RESULT 13  
GSPD\_ERWCH  
ID GSPD\_ERWCH STANDARD; PRT; 710 AA.  
AC Q01565;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)

01-NOV-1995 (Rel. 32, Last annotation update)  
General secretion pathway protein D precursor (Pectic enzymes  
secretion protein outd).  
OUTD.  
OS Erwinia chrysanthemi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Pectobacterium.  
OX NCBI\_TaxID=556;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=9337; PubMed=1453958;  
RX MEDLINE=93086427; PubMed=1453958;  
RA Condemine G., Dorel C., Hugouvieux-Cotte-Pattat N., Robert-Baudouy J.;  
RT "Some of the out genes involved in the secretion of pectate lyases in  
Erwinia chrysanthemi are regulated by kgR.";  
RL Mol. Microbiol. 6:3199-3211(1992).  
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE  
EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLOCATION OF THE  
MULTIPLE PECTIC ENZYMES.  
CC -!- SUBCELLULAR LOCATION: Outer membrane (Probable).  
CC -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X65265; CAA46370.1; -;  
DR PIR; S28014; S28014.  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSP1/IIIProtein.  
DR InterPro; IPR004845; GSP1/IIIProteinC.  
DR InterPro; IPR005644; NOLW-like.  
DR Pfam; PF00263; GSP1\_III; 1.  
DR Pfam; PF03958; GSP1\_III; 3.  
DR PRINTS; PR00811; BCTERIALGSPD.  
DR PROSITE; PS00875; T2SP.D; 1.  
KW Transport; Outer membrane; Signal.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 710 GENERAL SECRETION PATHWAY PROTEIN D.  
FT DOMAIN 288 353 GLY/SER-RICH.  
FT SEQUENCE 710 AA; 76213 MW; 156E84CC50C54FA CRC64;  
-----  
Query Match 8.5%; Score 328; DB 1; Length 710;  
Best Local Similarity 27.4%; Pred. No. 2, 1e-11;  
Matches 85; Conservative 77; Mismatches 134; Indels 14; Gaps 7;  
-----  
QY 464 NADTTGNRLTVSGRG-SYLDIPATNLIVTDRSVIEKFKRLDELDPVQACQWMEARI 522  
DB 347 STSSGGRTITIQGKEVTVAHDQTNLSLITAPPDMDRLDQVINDLRPQVLVEAII 406  
-----  
QY 523 VEAADGFSRLGKTCATGKXKXKNDTSAGFGVNSGFGDDKGAETKINLPITAAANS 582  
DB 407 AEIQADGUNGLOW--ANKRAGMTQFTNGIPSTAVIGTDQFRSNGTLT---TAYASA 461  
-----  
QY 583 ISLVRAISSGAL----NLELSASELSKTKTLIANPRVLTKRKEAKIESGVIEP-FTVTS 637  
DB 462 LSSFNQVTAQFYRGNWSMLLTALSSDSKNDVLATPSIVTLDNMEATFNVGQEVPLTGSQ 521  
-----  
QY 638 IANGSSSTTELKXAV-LGLTTPNITPDQIIMTVKINKDSPAQCSNQNTILCI--ST 694  
DB 522 TTSADNIFNTVERKTVGKLRKVPQNEGDSVLLQIEQVSVVADSNSTSSGLGVNT 581  
-----  
QY 695 KNLNTQAMVNGGTLIVGGIYEEDNGNTLTKVPLLDGIPIVGNLTKRGTDRRELLIF 754  
DB 582 RTVNNAVMTNGETVVVVGGLDKTSVESNDKVPVLLGDIPLWGLSFRSKSQEVRKXNMLF 641  
-----  
QY 755 ITPRINGTAG 764  
DB 642 LRPTIIRDFG 651

RESULT 14  
GSPD\_VIBCH STANDARD; PRT; 674 AA.  
ID AC F45779;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE General secretion pathway protein D precursor (Cholera toxin secretion  
protein epad).  
DE protein epad.  
GN EPSD OR VC2733.  
OS Vibrionaceae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=EJ Tor TRH7000;  
RA Overbye L.J.;  
RT "Organization of the general secretion pathway genes in Vibrio  
cholerae.";  
RL Thesis (1994), Michigan State University / East Lansing, U.S.A.  
[2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=EJ Tor N16961 / Serotype O1;  
RX MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
RA McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,  
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
cholerae.";  
RL Nature 406:477-483(2000).  
CC -!- FUNCTION: REQUIRED FOR SECRETION OF CHOLERA TOXIN THROUGH THE  
OUTER MEMBRANE.  
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).  
CC -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.  
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
FRAMESHIFT (POSSIBLY NATURAL) IN POSITION 55.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; L33796; AAA58785.1; -;  
DR EMBL; AE004338; -; NOT\_ANNOTATED\_CDS.  
DR TIGR; VC2733; -;  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSP1/IIIProtein.  
DR InterPro; IPR004845; GSP1/IIIProteinC.  
DR InterPro; IPR005644; NOLW-like.  
DR Pfam; PF00263; GSP1\_III; 1.  
DR Pfam; PF03958; GSP1\_III; 3.  
DR PRINTS; PR00811; BCTERIALGSPD.  
DR PROSITE; PS00875; T2SP.D; 1.  
KW Transport; Outer membrane; Signal; Complete proteome.  
FT SIGNAL 1 24 POTENTIAL.  
FT CHAIN 25 674 GENERAL SECRETION PATHWAY PROTEIN D.  
FT CONFLICT 89 89 V -> A (IN REF. 1).  
FT CONFLICT 144 144 R -> P (IN REF. 1).  
SQ SEQUENCE 674 AA; 73469 MW; 3D77B891A59E6223 CRC64;  
-----  
Query Match 8.0%; Score 309.5; DB 1; Length 674;  
Best Local Similarity 23.5%; Pred. No. 2, 2e-10;  
Matches 135; Conservative 111; Mismatches 216; Indels 113; Gaps 22;

QY 232 KQGNAGIIELALGAGQPDISQOHDHIIIVTKN---HTLPTTLQSLDADFKTPVQK 288  
DB 103 KDAKTSALPVLSEERANGDEVITQ-----VVAVNVSVRELSPLQLQIDNAGAGNVV-- 156  
QY 289 VTLKRLNDTQLIITTAGNWEVYKSAAGYFFQVLPK-----KONLESGGVNNAKPTFT 343  
DB 157 -----HYDPANILITGRAAVNRLA-----EIRRVDOAGDEKEIEVVELNNA----- 199  
QY 344 TGRKISLDFODVEIRITLQILAKESGNVIVASDSVNGKMTLSLKDVPWQALDLVWQARN 403  
DB 200 -----SAAEWRIVEALNK-----TTDAQNTPEFLKPKFV----- 229  
QY 404 LDMRQOQNIYNIAP--RDELLAKDKAPLQAEKDIALGALYQNFOLKYNVVEEPRSLIR 461  
DB 230 ADERTNSILISGDPKVRERL---KRLIKQLDVEWAAGK--NNEVWYLYKAKABDLVEVLK 284  
QY 462 -----LDNADTTGNRLTVSGSGVLI--DPANTLITVDTSRVIEKFKLIDELDVPAQ 514  
DB 285 GVSENLOAEKGTQPTT--SKRNEVMIHAHADNSLVLTAPQDIMANLEVICQLDIRRA 342  
QY 515 QVMIEARIVEAAGDFSRDLGKFEKA-----TGKKKLKNDTSAGF-- 553  
DB 343 QVLIEALIVEMAEQDGINLGQVWGSLESQSVIOVGNLTGASIGNVMIGLEAKDTTQTKAV 402  
QY 554 WGVNSGGGDDKWAETKINLPITAAANSISLVR-----AISSGALNLELSASELSKTKT 609  
DB 403 YDTNNFNRLNE---TTTTKGYTKLASUSSIQGAASVSTAMGDWTALINAVENDSSNI 458  
QY 610 LANPRVLTQNRKEAKIESGYEIPFTVTSIANGSGSTN---TELKAVLGLTVPNITPD 665  
DB 459 LSPSITVMDNGEASFIVGEVPP-VITGSTAGSNDNPNFQTVDRKEVGKIKVVPQINEG 517  
QY 666 GQIIMTVKINKDSPAQASGNQITLCISTKNLNTQAMVNGGTLIVGGIYEEDNGNLTG 725  
DB 518 NSVOLNIE--QEVSNVLGANGAVDR-FAKQLNTISVWQDQOMVLGLGULDERALESEK 575  
QY 726 VPLIGDIPVIGNLFPKTRGKKTDRRELLIFITPRIM 760  
DB 576 VPLIGDIPLLGQPRSTSSQVEKKNLVFIKPTII 610

RESULT 15  
VG4\_BP122 STANDARD; PRT; 428 AA.  
ID\_VG4\_BP122 AC P15420;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-JUL-1993 (Rel. 26, Last annotation update)  
DE Gene IV protein (GP1V).  
GN Bacteriophage I2-2.  
OS Viruses; ssDNA viruses; Inoviridae; Inovirus.  
OX NCBI\_TaxID=10869;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schoenmakers H.F.P.M., Yu M., Konings R.N.H.;  
RL Submitted (FEB-1989) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: THE EXACT FUNCTION OF THIS PROTEIN IS UNKNOWN ALTHOUGH  
CC IT MAY BE INVOLVED IN PHASE ASSEMBLY.  
CC -!- SIMILARITY: BELONGS TO THE PULD/OUTD/EXED/XPSD FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X14336; CAA32520.1; -  
DR PIR; S08084; S08084.  
DR InterPro; IPR001775; Bac\_GSPD.

DR InterPro; IPR004846; GSP11/IIIprotein.  
DR InterPro; IPR004845; GSP11proteinc.  
DR InterPro; IPR005644; Nolv-like.  
DR Pfam; PF00263; GSP11\_III; 1.  
DR Pfam; PF03958; GSP11\_III\_N.1.  
DR PRINTS; PR00811; BCTERIALGSPD.  
DR PROSITE; PS00875; T2SP\_D; 1.  
KW Phase maturation.  
SQ SEQUENCE 428 AA; 45252 MW; DA0A4E9103A38A42 CRC64;  
Query Match 8.0%; Score 309; DB 1; Length 428;  
Best Local Similarity 22.8%; Pred. No. 1.3e-10;  
Matches 104; Conservative 85; Mismatches 179; Indels 88; Gaps 14;  
QY 336 VNNAPKTTTGRKISLDFQDVEIRITLQILAKESGNVIVASDSVNGKMTLSLKD--PMD 392  
DB 27 LNNSP-----VRSEVQWYSKTKGKSVIWNPDVKGNITVFNADVNNANID 70  
QY 393 QALDVMQARNLDRQOQNIYNIAPRDELLAKDKAFLQAEKDIALGALYS----- 443  
DB 71 DFFKSVLANGL-VVAGNPVAVVSTPLTKLASQPSNEETYYDDES-GVAYEAVPOSAAAPA 128  
QY 444 -----QNFQKLYKNVEEPRS--ILRLDNADTTGNRLTVSGSGVLIDPATNLTIVTD 494  
DB 129 VPADLTVERNF-----NVTRVRSSDVLPLAKIFVDSN-----GGGNVVDYPGNNLSLVSG 177  
QY 495 TRSVIEKFRKLIDELDPVPAQVMIPEARIVE--AADFSRDLGVKFGATGKKLKNDSAF 552  
DB 178 SAQVNPALSDFTISDVAREQVLIQSLMPFETSVNGVDLSFALASGKVGAGGNTSAL 237  
QY 553 GGVNSGGGDDKWAETKINLPITAAANSISLVRASSGALNLELSASELSKTKTLAN 612  
DB 238 G-----TALSTAGSGFI---FNGNIALSLQAVQSDSNSKVIST 274  
QY 613 PRLVLTQNRKEAKIESGYEIPFTVTSIANGSGSTNT-----ELKAVLGLTVPNITPDQO 667  
DB 275 PRLTQSGQSGVSYGVQNVFPVTVGTGTEAASVNNPFTIERRDVGVSLKVTVPVNGNQ 334  
QY 668 IIMTVKINKDSPAQASGNQIT---LCISTKNLNTQAMVNGGTLIVGGIYEEDNGNLT 724  
DB 335 LVLTITDKADS-----LSNQAIASDIITNQRIQTIVIKDQITLLGLLSSNQFDSOR 389  
QY 725 KVPILGDIPIVIGNLFPKTRGKKTDRRELLIFITPRIM 760  
DB 390 SVFFNSKIPLIGLWLFRESHSDSKDDRTMFVLTAAHVI 425

RESULT 16  
GSPD\_XANCP STANDARD; PRT; 759 AA.  
ID\_GSPD\_XANCP AC P29041; P31763;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE General secretion pathway protein D precursor.  
GN XPSD OR PEFD OR XCC0670.  
OS Xanthomonas campestris (pv. campestris).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xanthomonas.  
OX NCBI\_TaxID=340;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92210513; PubMed=1313415;  
RA Hu N.-T., Hung M.-N., Chlou S.-J., Tang P., Chiang D.-C., Huang H.-Y.,  
RA Wu C.-Y.;  
RT "Cloning and characterization of a gene required for the secretion of  
RT extracellular enzymes across the outer membrane by Xanthomonas  
RT campestris pv. campestris."  
RL J. Bacteriol. 174:2679-2687(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=ATCC 33913 / NCPPB 528;  
RX MEDLINE=22022145; PubMed=12024217;







```

557 NSFGGDDKWAETKINLPITAAANSISLVRAISSGALNLELSASESLSKTKTLANPVL 616
231 NT-----DLTSLVSSAGGSFGI---FNGDVLGSLVRAKTNHSHKILSVPRIL 276
617 TONRKEAKIESGYEIPFTVTSIANGGSSTNT-----ELKAVLGLTVTNIPTDQGIIMT 671
277 TLSGKGKGISVGQNVPIITGRVTGESANVNPQTIERQVNGISMSVFPVAMAGNIVLD 336
672 VKINKDSFAQCASGNQITLCISTKVLNTQAMVENGGLTVGGIYEEDNGNTLTKVPLIGD 731
337 ITSADKSLSSSTQASDVI--TNQRSIATTVNLRDGQTLGLGLTDYKNTSQDSQGVFFLSK 394
732 IPIVGNLFKTRGKTKTDRELLIFITPRIM 760
395 IFLIGLLFSSRSDSNEESTLVLVKATIV 423

RESULT 19
VG4_BFPD
ID VG4_BFPD STANDARD; PRT; 426 AA.
AC P03564;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE Gene IV protein (gpiv).
GN IV.
OS Bacteriophage fd.
OC Viruses; ssDNA viruses; Inoviridae; Inovirus.
NCBI_TaxID=10864;
[1]
SEQUENCE FROM N.A.
STRAN=478 / Heidelberg;
MEDLINE=79136480; PubMed=745987;
Beck E., Sommer R., Auerswald E.A., Kurz C., Zink B., Osterburg G.,
Schaller H., Sugimoto K., Sugisaki H., Okamoto T., Takanami M.;
RA "Nucleotide sequence of bacteriophage fd DNA.";
RT Nucleic Acids Res. 5:4495-4503(1978).
CC -! FUNCTION: THE EXACT FUNCTION OF THIS PROTEIN IS UNKNOWN ALTHOUGH
CC IT MAY BE INVOLVED IN PHASE ASSEMBLY.
CC -! SIMILARITY: BELONGS TO THE PULD/OUTD/EXED/XPDS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; V00602; CAA23854.1; -.
CC FMBL; J02451; AAA32312.1; -.
CC PIR; A04268; Z4BFPD.
CC InterPro; IPR001775; Bac_GSPD.
CC InterPro; IPR004846; GSPII/I1protein.
CC InterPro; IPR004845; GSPIIproteinC.
CC InterPro; IPR005644; NOLW-like.
CC Pfam; PF00263; GSPII_I1; 1.
CC Pfam; PF03958; GSPII_III_N; 1.
CC PRINTS; PR00811; BCTERIALGSPD.
CC PROSITE; PS00875; T2SP_D; 1.
CC Phage maturation.
CC SEQUENCE 426 AA; 45822 MW; 5554766B843F7CB CRC64;

Query Match 7.6%; Score 292.5; DB 1; Length 426;
Best Local Similarity 20.7%; Pred. No. 1.1e-09;
Matches 93; Conservative 96; Mismatches 185; Indels 75; Gaps 12

Qy 336 VNNAPKFTTGRKISLDFQDVEIRITLQILAKESGMNIVASDSVNGKMTLSLKDPVWDQAL 395
Db 26 MNNSP-----LRDFVTWYSQTSGESVIVSDVKVGTIVTVSSDVKPENUR 69
Qy 396 DL---VMQARNLDMRQQGNVNIAPR-----DELLAKDKAFLOAEKDIAIDGLAY- 442

```

Db 70 NFFISVL RANFDM--VGSIPSI IQKNPN SQDI D E L P S S D--IQEYDDNSAPSGGFFV 126  
 Qy 443 -----SQNFOLKYNKVNVEERSILRLDNADTTGNRNVLVSGRGSVLIDPATNTLIVTDR 496  
 Db 126 PQNDNVITQT KINNVRKADLRVVELFVKSNKTSKSNVLSVDPGS-----NLLVVSAPK 178  
 Qy 497 SVIEKFKRLIDELVPAQOQVMIEARIEAADGPSRDLGVKFGATGKKLKNDTSAFGWGV 556  
 Db 179 DILNLPQFLSTVDLPFDQILIEGLIFEVQOGDALDFSFAAG-----SQRGTVAGCV 230  
 Qy 557 NSGFGGDDKGAETKINLPITAAANSISLVRALSISALNLELSASELSKTKYTLANPRVL 616  
 Db 231 NT-----DRLTSVLSAGGSGFGI---FNGDVLGLSVRAALKTNHSHLSVPRIL 276  
 Qy 617 TONRKEAKIESGYBIPTVTSIANGSGSSTN-----TELKKAIVLGLTVTNIITPDQIIMT 671  
 Db 277 TLSGQKSGISVGQNVPIITGRVTGESANVNNPQTVERQVNGVISMVSFFVAMAGNIVLD 336  
 Qy 672 VKINKSPAQACASGNQITLCISKTUNLTQAMVNGGTLVGGIYEDNGNTLTKVPLIGD 731  
 Db 337 ITS KADSLSSSTQASDVI--TNORSATTNLRDGTLLGLGLTDYKNTSQDSGVFFLSK 394  
 Qy 732 IPVTGNLFKTRGKKTDRRELLIFITPRIM 760  
 Db 395 IPLGLLFSSRSDSNEESTLYVLVKATIV 423  
 RESULT 20  
 ID\_VG4\_BP1F1 STANDARD; PRF: 429 AA.  
 AC 080300;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Gene IV protein (GP1V).  
 GN 4 OR IV.  
 OS Bacteriophage Iphi.  
 OC Viruses; ssDNA viruses; Inoviridae; Inovirus.  
 OX NCBI\_TaxID=10868;  
 [1]\_\_\_\_\_  
 KP SEQUENCE FROM N.A.  
 RA "Hill D.F., Hughes G., McNaughton J.C., Stockwell P.A., Petersen G.B.;  
 RT "DNA sequence of the filamentous coliphage Iphi".  
 RL Submitted (Oct-1993) to the EMBL/GenBank/DBJ databases.  
 CC -! FUNCTION: THE EXACT FUNCTION OF THIS PROTEIN IS UNKNOWN ALTHOUGH  
 CC IT MAY BE INVOLVED IN PHAGE ASSEMBLY (BY SIMILARITY).  
 CC -! SIMILARITY: BELONGS TO THE PULD/OUTD/EXED/XSPD FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U02303; AAC62158.1;  
 DR InterPro; IPR001775; Bac\_GSPD.  
 DR InterPro; IPR004846; GSP1/IIIProtein.  
 DR InterPro; IPR004845; GSP1/proteinC.  
 DR InterPro; IPR005844; NoIw-like.  
 DR Pfam; PF00263; GSP1I\_III; 1.  
 DR Pfam; PF03958; GSP1I\_III\_N; 1.  
 DR PRINTS; PRO0811; BCTERIALGSPD.  
 DR PROSITE; PS00875; T2SP\_D; 1.  
 KW Phage maturation.  
 SQ SEQUENCE 429 AA; 45863 MW; 9170B3B75C3915C5 CRC64;  
 Query Watch 7.6%; Score 292.5; DB 1; Length 429;  
 Best Local Similarity 22.2%; Pred.No. 1.1e-09;  
 Matches 104; Conservative 84; Mismatches 176; Indels 105; Gaps 194  
 Qy 342 TPTGKRIKSLDPQDVEIRTIQLQAKESGMNIVASDSVNGKMTUSLKDQFWD---QALDIV 398

DB 13 SFQSAIIVELNNAVREFFVSWYKTKQVPIIPDPVKGELTVYSADVTKDELFPQFTSV 72  
QY 399 MQARNLDRQOQN--IVNIAPRDELLAKD-----KAPLQAE 432  
DB 73 LRANGFOL-SPGNPAVQVKFNKNTVEYSDSPSEVPASSYDQDVPPTGDFKPEIRAN 131  
QY 433 KDIADLGLYSONFOLKYKVEEFSILR--LDNADTTGNRTNLVSGRGSVLDPATNLT 490  
DB 132 -----LITQYVNNVRAKLDAPVIDIEKGNITAGTKVPEMGR-----IFL 174  
QY 491 IYVTRSVIEKRLIDELDPVQAQVMEIARIVE--AADGFRDLGVKFGATGKKLKN 548  
DB 175 LVTASASQKELAEAFPSVDVPTQVLVESVIFETASDGP----- 215  
QY 549 TSAGMGVNSGFGGDKGAETKINLPITAAANSISLVRASS-----GALNLEL 598  
DB 216 -----DFSAAGDPSS-----PVAGGINTRLTSVLSSTGSGFGIFNGNIGLSL 261  
QY 599 SASSELSKTKTLIANPRVLTQNRKEAKIESGYEIPPTVTSIANGSSNT-----ELKAV 653  
DB 262 KALETSKSTLLSPRIILTMSPQGTFTAGQNVPEVTVGTGEAANNVNPFTIERHDVG 321  
QY 654 LGLTVTNPITPDGQIIMTVKINKDSPACASQNTI--LCISTKNLNTQAMVNGGTLIV 711  
DB 322 ISLKVPVPTPGGLIMDVSTNADS-----ISDSQTASDIITNRSISTTVQLKSGQTVLL 377  
QY 712 GGIYEDNGNTLVKPLGDIPIVGNLTKRGKTKDRRELLIFITPRIM 760  
DB 378 GGMVDNRSDSSVPWYKIPGLIGALFTSKSSNANKETLILIRARVV 426

RESULT 21  
VG4\_BPIKE  
ID VG4\_BPIKE STANDARD; PRT; 437 AA.  
AC P03667;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Gene IV protein (GPIV).  
GN IV.  
OS Bacteriophage IKe.  
OC Viruses; ssDNA viruses; Inoviridae; Inovirus.  
OX NCBI\_TaxID=10867;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85160831; PubMed=3981635;  
RA Peeters B.P.H., Peters R.M., Schoenmakers J.G.G., Konings R.N.H.;  
RT "Nucleotide sequence and genetic organization of the genome of the N-  
specific filamentous bacteriophage IKe. Comparison with the genome of  
the P-specific filamentous phages M13, fd and f1";  
RL J. Mol. Biol. 181:27-39(1985).  
CC -!- FUNCTION: THE EXACT FUNCTION OF THIS PROTEIN IS UNKNOWN ALTHOUGH  
IT MAY BE INVOLVED IN PHAGE ASSEMBLY.  
CC -!- SIMILARITY: BELONGS TO THE PULD/OUTD/EXED/XPSD FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X02139; CAA26076.1; -  
DR PIR; A04269; 24BPIK.  
DR InterPro; IPR001775; Bac GSPD.  
DR InterPro; IPR004846; GSPFI/IIIprotein.  
DR InterPro; IPR004845; GSPFIprotein.  
DR InterPro; IPR005644; Nolv-like.  
DR Pfam; PF00263; GSPFI\_III; 1.  
DR Pfam; PF03958; GSPFI\_III N; 1.  
DR PRINTS; PR00811; BCTERIALGSPD.

DR PROSITE; PS00875; T2SP\_D; 1.  
SQ SEQUENCE 437 AA; 46485 MW; D2456EF357F0D819 CRC64;  
Query Match 7.5%; Score 290; DB 1; Length 437;  
Best Local Similarity 21.4%; Pred. No. 1.6e-09;  
Matches 98; Conservative 92; Mismatches 201; Indels 66; Gaps 11;  
QY 322 FOVLPPKKNLESQVNNAPKFTTGRKISLDFQDVEIRITLIQILAKESGMNIVASDSYNGK 381  
DB 26 FNLVADPNL-----NNAP-----VRSFVQWYSQKSKAVVNPVUKGN 64  
QY 382 MTLSLKDVPMQALDLVNOARNLD--MRQQGNIVNIAPRDELLA-----KDKAFLQ 430  
DB 65 ITFVNADVNQANIDIDFFKSVLNFANGFVLMAGDPGSGVSTPSKLPSCQTDNDDDYEDSDYV 124  
QY 431 AEKDIADLGLYSONFOLKYKVEEFSILRDLNADTTGNRTNLV--SGRGSVLDPATN 488  
DB 125 FVGDVSPVSAQPKQLDLTVNFX-----LTVRSSDVLPLAKIFVDSNGGSDVIDYFGNN 180  
QY 489 TLIVTDRSVIEKFKLIDELDPVQAQVMEIARIVEAADGFRDLGVKFGATGKKLKN 548  
DB 181 SLVSGSAIINALADFITSDIVARDQVLIQSLMFTSLVNGVDLSFAAG----- 230  
QY 549 TSAGMGVNSGFGGDKGAETKINLPITAAANSISLVRASSGALNLELSASELSKTK 608  
DB 231 -SASDKVAGFN-----TSALGTALSTAGSGFI---PNGNVLALSIOAVKNDNSK 279  
QY 609 TLIANPRVLTQNRKEAKIESGYEIPPTVTSIANGSSNT-----ELKAVLGLTVTPTNIT 663  
DB 280 VISTPRILTSQGTGYISVGQNVPEVTVGTGEAANNVNPFTIERDRDVGSLVKVTPVVM 339  
QY 664 PDGQIIMTVKINKDSPACASQNTILCISTKNLNTQAMVNGGTLIVGGIYEEDNGNTL 723  
DB 340 GNGQLVLIIDTKADSLTSQMTASDI--TNQRHQTTVQIKDQGTLLGLGLSDNTDGN 397  
QY 724 TKVPLGDIPIVGNLTKRGKTKDRRELLIFITPRIM 760  
DB 398 RSPWPFESVPVIGWLFGRSHSDSHNERTMFVLLTAHVI 434

RESULT 22  
YSCC\_YEREN  
ID YSCC\_YEREN STANDARD; PRT; 607 AA.  
AC Q01244;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE YCP proteins translocation protein C precursor.  
GN YSCC.  
OS Yersinia enterocolitica.  
OG Plasmid pYV.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersinia.  
OX NCBI\_TaxID=630;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=439-80 / Serotype O:9;  
RX MEDLINE=9131716; PubMed=1860816;  
RA Michiels T., Vanooteghem J.-C., de Rouvroit C., China B., Gustin A.,  
RA Boudry P., Cornelis G.R.;  
RT "Analysis of virC, an operon involved in the secretion of Yop  
proteins by Yersinia enterocolitica";  
RL J. Bacteriol. 173:4994-5009(1991).  
CC -!- FUNCTION: VERY LIKELY REQUIRED FOR THE EXPORT PROCESS OF THE YOP  
PROTEINS.  
CC -!- SUBCELLULAR LOCATION: Outer membrane (Probable).  
CC -!- SIMILARITY: BELONGS TO THE PULD/OUTD/EXED/XPSD FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M74011; AAC37020.1; -
DR PIR; C40361; C40361.
DR InterPro; IPR001775; Bac GSPD.
DR InterPro; IPR004846; GSPII/IIIprotein.
DR InterPro; IPR004845; GSPII/IIIproteinC.
DR InterPro; IPR005644; NoW-like.
DR InterPro; IPR003522; SecIII_OMP.
DR Pfam; PF00263; GSPII_III; 1.
DR Pfam; PF03958; GSPII_III_N; 2.
DR PRINTS; PR00811; BCTERIALGSPD.
DR PRINTS; PR01337; TYPE3OMGPROT.
DR PROSITE; PS00875; T2SP D; 1.
KW SIGNAL; Transport; protein transport; Outer membrane; Signal.
FT SIGNAL 26 POTENTIAL.
FT CHAIN 27 607 YOP PROTEINS TRANSLLOCATION PROTEIN C.
SQ SEQUENCE 607 AA; 67209 MW; CC5EA81348F3C687 CRC64;
Query Match 7.4%; Score 284.5; DB 1; Length 607;
Best Local Similarity 24.2%; Pred. No. 5.1e-09;
Matches 119; Conservative 74; Mismatches 185; Indels 113; Gaps 13;
QY 357 IRTILOILAKESGMIVASDVNGKMTLSLK-DVPHQDQALDLMQARNLDMRQGNVNI 415
DB 45 LRDLTDFGANYDATVVSCKINDKRVSGQFEHNPQD-FLQHIASLYNLVWYDGNLYI 103
QY 416 -----APRD-ELLAK 424
DB 104 FKNSEVASRLRLQSEAAELKQALQSGIWEPRFCWGRPDASNRLVYVSGPRYLELVEQ 163
QY 425 DKAFLOAKDIAD-LGALYSQNFQIKYKNVEE-----FRSILRLDNADTT 468
DB 164 TAAALEQQTQIRSEKTAIAEIFPLKYSASDRTIHYRDEVAAPGVATILQRVLSDAT 223
QY 469 GNENTL-----VSGRGSLVDPAINTLIIVTDRSVIEKFRKLIDELDVPAQQVM 517
DB 224 IQQVTVDNQRIPOAATRASQAARVEADPSLNAIIVRDSPEPMYQRLIHALDKPSARIE 283
QY 518 IEARIVEADGFRDLGVKFGATGKKLKNDSAFGCVNGSFGGDDKWAETKINLPIT 577
DB 284 VALSIVDINADQTELVG-----DMRVGIRGTGNHQQVVIKTGDQSNV 326
QY 578 AAANSI-SLVRAISSGALNLELSASELSKTKTLANPRVLTVQNRKEAKIESG--YEIPFT 634
DB 327 ASNGALGSLVDARGLDVLLARVNLLENESGAQVSRPTLLTQENAGAVIDHSETYVYKT 386
QY 635 VTSIANGSGSTNELKAVLG--LTVPNITPDG---QIMTVKINKDSPAQACSGNQTI 689
DB 387 GKEVA-----ELKGITYGTWLRTPRVLTQGDKSEISLNLHIEDGNQKPNSSGIEGI 438
QY 690 LCISTGNLNTQAMVNGGTLVGGIYEEDNGNTLTQVPLIGDIPVIGNLFKTRGKKTDRR 749
DB 439 PTISRTVVDIVARVGHQOSLIIGIYRDELVSVALSKVPLIGDIPYIGALFRKSELTRT 749
QY 750 ELLIFITPRIM 760
DB 499 VRLFIIEPRII 509
RESULT 23
ID HRP_H_PSESY STANDARD; PRT; 701 AA.
AC Q01723;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Hypersensitivity response secretion protein hrph precursor.
GN HRP_H.
OS Pseudomonas syringae (pv. syringae).
CC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
```

```
CC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=321;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=61;
RX MEDLINE=93015750; PubMed=1400238;
RA Huang H.-C., He S.-Y., Bauer D.W., Collmer A.;
RT "The Pseudomonas syringae pv. syringae 61 hrph product, an envelope
RT protein required for elicitation of the hypersensitive response in
RT plants.";
RL J. Bacteriol. 174:6878-6885(1992).
RN [2]
RP REVISIONS.
RA Deng W.-L.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-52 FROM N.A.
RC STRAIN=61;
RX MEDLINE=96025090; PubMed=7579617;
RA Huang H.-C., Lin R.H., Chang C.-J., Collmer A., Deng W.-L.;
RT "The complete hrp gene cluster of Pseudomonas syringae pv. syringae
RT 61 includes two blocks of genes required for harpin secretion that
RT are arranged collinearly with Yersinia ysc homologs.";
RL Mol. Plant Microbe Interact. 8:733-746(1995).
CC -!- FUNCTION: INVOLVED IN THE SECRETION OF A PROTEINACEOUS ELICITOR OF
CC THE HYPERSENSITIVITY RESPONSE IN PLANTS.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE PULD/OUTD/EXED/XPSD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; L01064; AAC05014.1; -
DR EMBL; U25813; AAC05085.1; -
DR InterPro; IPR001775; Bac GSPD.
DR InterPro; IPR004846; GSPII/IIIprotein.
DR InterPro; IPR004845; GSPII/IIIproteinC.
DR InterPro; IPR005644; NoW-like.
DR InterPro; IPR003522; SecIII_OMP.
DR Pfam; PF00263; GSPII_III; 1.
DR Pfam; PF03958; GSPII_III_N; 2.
DR PRINTS; PR00811; BCTERIALGSPD.
DR PRINTS; PR01337; TYPE3OMGPROT.
DR PROSITE; PS00875; T2SP D; FALSE NEG.
KW Protein transport; Transport; Signal; Outer membrane; Translocation;
KW Hypersensitive response.
FT SIGNAL 1 21
FT CHAIN 22 701 POTENTIAL.
FT CHAIN 22 701 HYPERSENSITIVITY RESPONSE SECRETION
FT CHAIN 22 701 PROTEIN HRP_H.
SQ SEQUENCE 701 AA; 76546 MW; 4F470B33B9D00025 CRC64;
Query Match 7.1%; Score 273.5; DB 1; Length 701;
Best Local Similarity 28.4%; Pred. No. 2.6e-08;
Matches 92; Conservative 49; Mismatches 131; Indels 52; Gaps 10;
QY 459 ILRLDNADTTGNRTLVSGRGSLVLI--DPAINTLIIVTDRSVIEKFRKLIDELDVPAQQV 516
DB 263 VLHYGGGCTKSGKSRSGGRANIVTADVRNNAVLIYDLPSRKAMYEKLIKELDVSRNLI 322
QY 517 MIEARIV-----EAADGFSRDLGVKFGATGKKLKNDSAFGCVNGSGF---GGDDKKGAE 569
DB 323 EIDAVILDIRNELAELSSR-----WNFNAGSVNGGAMFDAQ 360
QY 570 TKINLPITAAANSISLVRAISSGALNLELSASELSKTKTLANPRVLTVQNRKEAKIESGY 629
DB 361 TSSTLFIO-----NAGFAELHLENGSGASVIGNPSILTLENQPAVIDFSR 408
QY 630 EIPFTVTSIANGSGSTNELKAVLGTVTP-NITPDG--QIMTVKINKDSPAQACSGN 686
```

```

Db 409 TEYLATS---ERVANIEPIAGTSLQVTPRSLDHGKFPQLIVDI-EDQIDISDN 463
QY 687 QILICISTKNTLQAMVNGTLLVGGIYEEDNGNTLTKVPLGDIPIVGN-LKTRGKK 745
Db 464 DTQPSVTKGNVSTQAVIAHSGSLVIGGPHGLEANDKVKHVPVLGDIPIYIGKLLFQSRSE 523
QY 746 TORRELLIIFITPRIMGTAGNSLRY 769
Db 524 LSQRERLFLTPRELIGDQNPARY 547

RESULT 24
HRA1_XANCV STANDARD; PRT; 607 AA.
AC P80151;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Hypersensitivity response secretion protein hrp1 precursor.
GN HRP1.
OS Xanthomonas campestris (pv. vesicatoria).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=341;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=Isolate 75-3;
RX MEDLINE=93113007; PubMed=1472717;
RA Fenselau S., Balbo I., Bonas U.;
RT "Determinants of pathogenicity in Xanthomonas campestris pv. vesicatoria are related to proteins involved in secretion in bacterial pathogens of animals.";
RL Mol. Plant Microbe Interact. 5:390-396(1992).
CC -!- FUNCTION: NECESSARY FOR BOTH BASIC PATHOGENICITY AND THE INDUCTION OF THE HYPERSENSITIVE RESPONSE IN RESISTANT PLANTS. COULD BE A PART OF A SPECIFIC TRANSPORT APPARATUS OR A SECRETION APPARATUS THAT IS REQUIRED FOR PATHOGENICITY. HRP PROTEINS MAY FORM A COMPLEX (TUNNEL/PORE) THAT ENABLES THE EXPORT OF MOLECULES SUCH AS VIRULENCE AND AVIRULENCE FACTORS.
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: BELONGS TO THE PULD/OUTD/EXED/XPSD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement.(See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M99173; AAA27603.1; -.
DR InterPro; IPR001775; Bac_GSPD.
DR InterPro; IPR004846; GSP1/IIIProtein.
DR InterPro; IPR004845; GSP1IproteinC.
DR InterPro; IPR005644; NslW-like.
DR InterPro; IPR003522; SecYII_OMP.
DR Pfam; PF00263; GSP1I_III; 1.
DR PRINTS; PR00811; BCTERIALGSPD.
DR PRINTS; PR01337; TYPE3OMGPROT.
DR PROSITE; PS00875; T2SP_D; 1.
KW Protein transport; Transpport; Signal; Outer membrane; Translocation;
KW Hypersensitive response.
FT SIGNAL 1 33
FT CHAIN 34 607
FT SEQUENCE 607 AA; 63935 MW; E6C6337C864F9308 CRC64;

Query Match 7.0%; Score 268; DB 1; Length 607;
Best Local Similarity 27.5%; Pred. No. 4.4e-06;
Matches 84; Conservative 56; Mismatches 141; Indels 24; Gaps 6;

```

```

QY 464 NADTTGNRNLVSGRGSVLIDPATNTLIVTDTSRVIEKFKRLIDELDVPAQOVMEARIV 523
Db 317 NPIDAGGAGELASDAPVIEADPRNTNAILIRDRPERMQSYGTLLIQOLDNRKLLQIDATII 376
QY 524 EAADGFSRDLGV--KFGATGKKLKNDSAFGMGWSGFGDDKWAETKINLPITAAAN 581
Db 377 BIRDGAMQDLGVDRWRHFSQHTDITGGRGOLGFGALSGAATDGATTPVGGTLTA--- 433
QY 582 SISLVRAISSGALNLELSASLSKTKTLANPRVLQNRKEAKIESGYEIPFTVTSIANG 641
Db 434 ---VLGDAGRYLMTRVSALETTNKAKIVSSPQVATLDNVEAVNDHKQCAFVRV---SG 485
QY 642 GSSTNTELKAVLGLTVTPNI---TPDGOIMTVKINKDSPAQCASGNQ---ILCISTK 695
Db 486 YASADLYNLSAGVSLRVLPSPVPGSPNGQKRLDRIEDGQ-----LGSNTVDGIPVITSS 540
QY 696 NLNTQAMVNGTLLVGGIYEEDNGNTLTKVPLGDIPIVGNLTKRGKTKDRRELIIFI 755
Db 541 EITTAQFVNGESLLIAGYAYDADETDLNATPGLSKIPLVGNLFKHKQSKGTRMORFL 600
QY 756 TPRIM 760
Db 601 TPHVV 605

RESULT 25
HRA1_RALSO STANDARD; PRT; 568 AA.
AC Q52498;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypersensitivity response secretion protein hrpA precursor.
GN HRP A OR HRCC OR RSP0874 OR RS01645.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=93113006; PubMed=1472716;
RA Gough C.L., Genin S., Zischek C., Boucher C.A.;
RT "hrp genes of Pseudomonas solanacearum are homologous to pathogenicity determinants of animal pathogenic bacteria and are conserved among plant pathogenic bacteria.";
RL Mol. Plant Microbe Interact. 5:384-389(1992).
RN [2];
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=95349395; PubMed=7623665;
RA van Gijsegem F., Gough C.L., Zischek C., Niqueux E., Arlat M., Genin S., Barberis P., German S., Castello P., Boucher C.A.;
RT "The hrp gene locus of Pseudomonas solanacearum, which controls the production of a type III secretion system, encodes eight proteins related to components of the bacterial flagellar biogenesis complex.";
RL Mol. Microbiol. 15:1095-1114(1995).
RN [3];
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S., Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L., Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N., Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T., Sigulier P., Thebault P., Whalen M., Wincker P., Levy M., Weissbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
CC -!- FUNCTION: INVOLVED IN THE SECRETION OF A PROTEINACEOUS ELICITOR OF THE HYPERSENSITIVITY RESPONSE IN PLANTS (BY SIMILARITY).

```



```
Db 262 EFRFGVSLFVPTVLDNLIHVKPEVSELSLOGAVQVNGIAPVATRRADTVVELA 321
Qy 705 NGCTLIVGGIYEDNGNTLTKVPLLDGIPVIGNLFTKTRKKKTDRELLIFITPRINGTAG 764
Db 322 SQGSFVIGGLIRNVNNDISAFPLWGRIPILGALFRSSSFQKSESLVILVTPYIVRPGS 381
Qy 765 N 765
Db 382 N 382

RESULT 28
MXID SHISO
ID MXID SHISO STANDARD; PRT; 566 AA.
AC Q55293; 1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Outer membrane protein mxid precursor.
GN MXID.
OS Shigella sonnei.
OG Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=624;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HM383;
RA Atakawa E., Kato J.I., Ito K.I., Watanabe H.;
RT "Comparison and high conservation of nucleotide sequences of spa-mxi
RT regions between S.sonnei and S.flexneri -- identification of a new
RT gene coding plausible membrane protein.";
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NECESSARY FOR THE SECRETION OF IPA INVASINS.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE PULD/OUTD/EXED/XPSD FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC -----
CC EMBL; D50601; BAA09154.1; -
CC InterPro; IPR001775; Bac_GSPD.
CC InterPro; IPR004846; GSP1/IIIProtein.
CC InterPro; IPR004845; GSP1IProteinC.
CC InterPro; IPR005644; Nclw-like.
CC InterPro; IPR003522; Sec111_OMEF.
CC Pfam; PF00263; GSP11_III; 1.
CC Pfam; PF03958; GSP11_III_N; 2.
CC PRINTS; PR00811; BCTERIALGSPD.
CC PRINTS; PR01337; TYPE3OMGPROT.
CC PROSITE; PS00875; T2SP.D; 1.
CC Virulence; Plasmid; Outer membrane; Transport; Protein transport;
KW Signal.
FT SIGNAL.
FT CHAIN.
SQ SEQUENCE 566 AA; 63144 MW; F9B0CE710DE4553D CRC64;

Query Match 5.5%; Score 213.5; DB 1; Length 566;
Best Local Similarity 24.7%; Pred. No. 5.1e-05;
Matches 71; Conservative 52; Mismatches 123; Indels 41; Gaps 9;

Qy 480 SVLIDPANTLVITRSVIEKRLIDELVPAQQVMEARIVEAADFSFDLGKVEGA 539
Db 265 SLIAYFETNSILVKNQDQIQIIRDIITOLDVAKRHIELSWIDKSELNILGV---- 320
Qy 540 TGKKLKNDSAFGQVNSFGGDDKGAETKINLPITAAANSISLVRAISSGALNELS 599
```

```
Db 321 -----NWQGTASFG--DSFGA--SFMSSASISLTDGNKFAS-----VM 357
Qy 600 ASELSLTKTLANPRVLTQNRKEAKIESGYBIPFTVTSIANGSGSTNTELKAVLG--LT 657
Db 358 ALNKKKANVSRPVLTQENI PAIFDNNR--TFVSLV-----GERNSSLHVTYGLIN 411
Qy 658 VTPNITPGQIIMTVKI-----NKDSPAQCASGQTILC-ISTKNLNTQAVVNGTGLIV 711
Db 412 VIPRESSRGOIEMSLTIEDGTGNSQSNYNNTSVLPVGRTKISTIAVPQGSLLI 471
Qy 712 GGIYEDNGNTLTKVPLLDGIPVIGNLFTKTRKKKTDRELLIFITPR 758
Db 472 GGYTHETNSNEIVSIPFLSSIPVIGNVFKYKTSNISNIVRVFLIQPR 518

RESULT 28
MXID SHIFL
ID MXID SHIFL STANDARD; PRT; 566 AA.
AC Q0461; Q8VSH0;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Outer membrane protein mxid precursor.
GN MXID OR CP0145.
OS Shigella flexneri.
OG Plasmid pWR100, and plasmid pCP301.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M90T / Serotype 5; PLASMID=pWR100;
RX MEDLINE=93172961; PubMed=8437520;
RA Allouai A., Sansonetti P.J., Parsot C.;
RT "MxiD, an outer membrane protein necessary for the secretion of the
RT Shigella flexneri lpa invasins.";
RL Mol. Microbiol. 7:59-68(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=M90T / Serotype 5; PLASMID=pWR100;
RX MEDLINE=20566792; PubMed=1115111.
RA Buchrieser C., Glaser P., Rusnick C., Nedjari H., d'Hauteville H.,
RA Kunst F., Sansonetti P., Parsot C.;
RT "The virulence plasmid pWR100 and the repertoire of proteins secreted
RT by the type III secretion apparatus of Shigella flexneri.";
RL Mol. Microbiol. 38:760-771(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=M90T / Serotype 5; PLASMID=pWR100;
RX MEDLINE=21189246; PubMed=11292750;
RA Venkatesan M.M., Goldberg M.B., Rose D.J., Grotbeck E.J., Burland V.,
RA Blattner F.R.;
RT "Complete DNA sequence and analysis of the large virulence plasmid of
RT Shigella flexneri.";
RL Infect. Immun. 69:3271-3285(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a; PLASMID=pCP301;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin O., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
CC -!- FUNCTION: NECESSARY FOR THE SECRETION OF IPA INVASINS.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE PULD/OUTD/EXED/XPSD FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
CC EMBL; X67206; CAA47644.1; --  
CC EMBL; AL391753; CAC05920.1; --  
CC EMBL; AF348706; AAL8464.1; --  
CC EMBL; AF386526; AAL72331.1; --  
CC PIR; S28068; S28068.  
CC InterPro; IPR001775; Bac\_GSPD.  
CC InterPro; IPR004846; GSPII/IiIprotein.  
CC InterPro; IPR004845; GSPiIproteinC.  
CC InterPro; IPR005644; NolW-like.  
CC InterPro; IPR003522; SeciIIi\_OMPg.  
CC Pfam; PF00263; GSPiI\_III; 1.  
CC Pfam; PF03958; GSPiI\_III\_N; 2.  
CC PRINTS; PR00611; BCTERIALGSPD.  
CC PRINTS; PR01337; TYPE3OMGPROT.  
CC PROSITE; PS00875; T2SP\_D; 1.  
KW Virulence; Plasmid; Outer membrane; Transport; Protein transport;  
Signal.  
KW Signal.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 566 OUTER MEMBRANE PROTEIN MXID.  
FT VARIANT 296 296 V--I (IN PLASMID PCP301).  
FT SEQUENCE 566 AA; 63172 MW; 5680D80B16C2431 CRC64;  
SQ  
  
Query Match 5.5%; Score 212.5; DB 1; Length 566;  
Best Local Similarity 24.7%; Pred. No. 5.8e-05;  
Matches 71; Conservative 52; Mismatches 123; Indels 41; Gaps 9;  
  
QY 480 SVLIDPATNTLVITDRSVIEKFKLIDELDPVPAQVMIARIVEAADGFSRDLGVKFGA 539  
DB 265 SLIATPENTSLVKNQNDQOIIRDIITOLDVAKRHIELSLWIIDKSELNNLGV--- 320  
QY 540 TGKKKLKNDTSAFGWGVNSFGGDDKGAETKINLPITAAANSISLVRAISSGALNLELS 599  
DB 321 -----NQWQTASFG--DSEGA--SFNMSSASISTLDGNKFIA-----VM 357  
QY 600 ASELSKTKTLANPRVLITQNRKEAKIESGYEIPFTVTSIANGSGSTNTELKAVLG--LT 657  
DB 358 ALNOKKANVVSRRPVILTQENIPALFDNNR--TFVSVL---GERNSSLEHVTYGTLIN 411  
QY 658 VTPNITPDGQIIMTVKI-----NKDSPAQASGNTQILC-ISTKNLTQAMVENGTLIV 711  
DB 412 VIPFSRSGQIEMSLTIEDGTGNSQSNYNUNENTSVLPEVGRTKISTIARVPOGKSLI 471  
QY 712 GGIVEEDNGNTLTVPKLLGDIPVLCNLFKTRGKKTDRRELLIFTPR 758  
DB 472 GGYTHETNSNELISIPFLSSIPVIGNVFKYKTSNISLVNRYFLIQPR 518  
  
RESULT 29  
INVG\_SALTY STANDARD; PRT; 562 AA.  
AC P35672;  
DT 01-JUN-1994 (Rel. 29, Created).  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE INVG protein precursor.  
GN INVG OR STM2898.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TMU;  
RX MEDLINE=95172719; PubMed=7869245;  
RA Lodge J., Douce G.R., Amin I.I., Bolton A.J., Martin G.D.,  
RA Chatfield S., Dougan G., Brown N.L., Stephen J.;  
RT "Biological and genetic characterization of TnpHoA mutants of

RT Salmonella typhimurium TWL in the context of gastroenteritis.";

RL Infect. Immun. 63:762-769(1995).

RL [2]

RL SEQUENCE FROM N.A.

RP STRAIN=SR-11;

RP MEDLINE=95089692; PubMed=7997169;

RA Kariga K., Bossio J.C., Galan J.E.;

RA "The Salmonella typhimurium invasion genes invF and invG encode

RT homologues of the AraC and Pula family of proteins.";

RL Mol. Microbiol. 13:555-568(1994).

RL [3]

RP SEQUENCE FROM N.A.

RP STRAIN=LT2 / SGSC412 / ATCC 700720;

RP MEDLINE=21534948; PubMed=11677609;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,

RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,

RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,

RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,

RA Waterston R., Wilson R.K.;

RT "Complete genome sequence of Salmonella enterica serovar Typhimurium

RL Lt2.";

RL Nature 413:852-856(2001).

RL H- FUNCTION: INVOLVED IN THE INVASION OF THE CELLS OF THE INTESTINAL

CC EPITHELIUM. COULD BE NECESSARY FOR THE EXPORT OF INVASION RELATED

CC DETERMINANTS.

CC -1- SUBCELLULAR LOCATION: Outer membrane (Potential).

CC -1- SIMILARITY: BELONGS TO THE PULB/OUTD/EXED/XPSD FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC at the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X75302; CAAS3049.1; -;

DR EMBL; U08280; AAA74040.1; -;

DR EMBL; AE008832; AAL2178.1; -;

DR PIR; S54420; S54420.

DR StyGene; SG10188; invG.

DR InterPro; IPR001775; Bac GSPD.

DR InterPro; IPR004845; GSPH/IIIprotein.

DR InterPro; IPR004845; GSPHproteinC.

DR InterPro; IPR005644; NoliW-like.

DR InterPro; IPR003522; SecIII\_OMPG.

DR Pfam; PF00263; GSPH\_III; 1.

DR Pfam; PF03958; GSPH\_III N; 2.

DR PRINTS; PR00811; BACTERIALGSPD.

DR PRINTS; PR01337; TYPE3OMGSPROT.

DR PROSITE; PS00875; T2SP\_D; 1.

DR Virulence; Transport; Protein transport; Signal; Outer membrane;

KW complete proteome.

KW SIGNAL 1 14

FT CHAIN 15 562

FT CONFLICT 12 12

FT CONFLICT 121 121

FT CONFLICT 197 205

FT CONFLICT 232 240

FT CONFLICT 243 243

FT CONFLICT 262 264

FT CONFLICT 328 328

FT CONFLICT 329 329

FT CONFLICT 370 380

FT CONFLICT 562 AA; 61765 MW; 8022905BE256058D CRC64;

SQ SEQUENCE

Query Match 4.6%; Score 178.5; DB 1; Length 562;

Best local similarity 20.4%; Pred. No. 0.0049;

Matches 125; Conservative

210 QTAAPAKQAAAAAPAKQTNIDFRDKGNAGIIEIA-ALG-----FAGQP-----DISQQDHDI 260

52 QLKPEVTVSQQAAARKKITGNFEFHPDNPALIEKLSLOGLTWYDPDGAIVYDASEMR-NA 110







Db 1228 NNASDAQPVTFVADKDSAVVVLQTSKAEIIGNGVDETTLTATVKDPFDNAVKDLQVTFST 1287  
Qy 322 ----FQVLPKQNLSESGVNAKPTFTGRKISLDLFDQDVEIRTILOILAKE-----SGMN-- 371  
Db 1288 NPADTQUSQSKNTNDGVA--AEVTEFG-----TVLGVHTAEATLPNGNDT 1332  
Qy 372 --IVASDSVNGKMTLSLKDVPMDQALDLVMQARNLDMRQ-----QGNIVNI 415  
Db 1333 KIVNIAPDASNAQVTL--NIPAOVV-----TNNSDSVQLTATVKDPSNHPVAGITVNF 1384  
Qy 416 APRDELLAKAFCAEKDADIGALYSQNFOLKYKNVEFRSILRLDNADTTGNEN-TL 474  
Db 1385 TPOQDVAAN-----FTLENNGIAITQANGAEHVTLKGGKAGHTHTVATLSNNNTSDSOPVTF 1441  
Qy 475 VSGRGSVL-----IDPATNTLIVIDTRSVIEKFKLIDELDV----- 511  
Db 1442 VADKTSALVWLOISKNEITGNGVDSALTATVKD-----QFDNEVNNLPVTFSTASSGL 1495  
Qy 512 ---PAQOVMEARIVEAADGSRDLGVKFG-ATCKKXKLNKDTSAFGHGVNSGFGDDKWG 567  
Db 1496 TLTGESNTNESHGIAQATLA-----GVAFGEQVTVASLANNGASDNKTVH--FIGD--- 1544  
Qy 568 AETKINLPITAAANSISLVRAISSGALNLSASELSKTKTLANPRVLTONRKEAKIES 627  
Db 1545 -----TAAAKIILTVPVDS-----LIAGTPONSSGSVITATVDNNGFPVK--- 1586  
Qy 628 GYEIPFT--VTSIANGGSSTNTELKAVLGLT-----VTPNITPP-----GOIM 670  
Db 1587 GVTYNFTSNAATAEMTNGQAVTNEQKATVYTNTRSSIESGARPPTVBEASLENGSSTL 1646  
Qy 671 TVKINKOSPACCA-----SGNOTILCISTKNLNTQAMVENGGLIV----- 711  
Db 1647 STSINVNADASTAHLTLLOALFDVTSAGDTNLTNLYEVKDNKNGVGPQOEVTLSVSGEV 1706  
Qy 712 -----GGIYEE-----DNGN-----TLTKVPLLGDI----- 732  
Db 1707 TPSNNAIYTNHNGNFYASFTATKAGYQVATLENGDSMQOQVTVYVPVNAEISLAAS 1766  
Qy 733 --PVIGNLFKTRGKTDRELLIIFITPRIMGTAGNSL 767  
Db 1767 KDPVIAN-----NNDLTTLTATVADTEGNAI 1792

RESULT 31  
ID\_TEGU\_HSVEB STANDARD; PRT; 3421 AA.  
AC P28955;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE Large tegument protein.  
GN 24.  
OS Equine herpesvirus type 1 (strain Abap) (EHV-1).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Varicellovirus.  
OX NCBI\_TaxID=31520;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92295566; PubMed=1318606;  
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;  
RT "The DNA sequence of equine herpesvirus-1";  
RL Virology 189:304-316(1992).  
CC -!- FUNCTION: TEGUMENT PROTEIN.  
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,  
CC EHV-1 24, EHV BFEL1, HVS-1 64, VZV 22, AND HCMV UL48.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----EMBL; M86664; AAB02459.1; --  
DR PIR; G36797; WZBES6  
DR InterPro; IPR006328; Herpes teg N.  
DR InterPro; IPR005210; Herpes\_UL36.  
DR Pfam; PF04843; Herpes\_teg\_N; 1.  
DR Pfam; PF03586; Herpes\_UL36; 1.  
SQ SEQUENCE 3421 AA; 367078 MW; 5075EFB4739BB7AC CRC64;  
  
Query Match 3.8%; Score 147; DB 1; Length 3421;  
Best Local Similarity 19.0%; Pred. No. 3.8;  
Matches 163; Conservative 88; Mismatches 279; Indels 328; Gaps 37;  
  
Qy 15 ATAATQATASAGNITDIKVSLLPNKQIKVY-SFDKEIVNPTGVTSTSSPARI-ALDFEOTG 72  
Db 2626 ASCATQSDSGKTLT-LDVEKTKQSKDKVVPVPTDK-----PS---TTTPAALKQSDASKPP 2677  
Qy 73 ISMDQVLEYADPLLSKISAAQNSRRARLNLN-----KPG----- 109  
Db 2678 TAAIQQKQKLGTPVTPKDSGDKPTONASAPVGVSVTPDGTGAKPPKPPDAPVDDTKQP 2737  
Qy 110 ---QYNTVEVRGNKWI-----FINESDDTVS-----APARPAVAAA 142  
Db 2738 VRKSLPSQVRGGRPVIRPSLGPFKFTGPGVTPVHGLPPSDSNVTQSTKEPPKPAVETP 2797  
Qy 143 PAAPAKQQQCRVTYQVRSIRIOTLYPGKTT---AAAPTESVVSUSAPSPAKQOAAAASA 199  
Db 2798 AAAPAKSAAAPAAAPAKSAAAPAAAPAKSAAAPAAAPAAAPAKSAAAPAAAPAAAPAA 2857  
Qy 200 KQCTAA-----PAKQOTAAAPAKQOAAAAPAKO-----TNIDFRKDGKXNAGHIELAA 244  
Db 2858 KQCTSAABVPKPAKQAKQAKQAKQAKQAKQAKQAKQAKQAKQAKQAKQAKQAKQAKQAK 2917  
Qy 245 LGFA-----QCPDISQOHDHIIIVTLKNHTLPTTLQKSL 277  
Db 2918 LSVPEKTLPLDPSFGAVPENTPLDPSFGSPDLS-----ASKNHTTDAVSSDRF 2970  
Qy 278 DVA-----DFKTPQVKVTLKRLND-----TQLIITTAGN----- 307  
Db 2971 SVACKVPLPDSDEDDFYSAVDVPLDPSPTDPPSSGRSDARAPTGVGVASIHKSDSRNN 3030  
Qy 308 ---WELV-----NKSA-----AFGYFTFQVLPKQNLSESG- 338  
Db 3031 ROSDAMRRAFADTLHGRPNRSATKPKSAPYKVPKHAISYTKIPVNPNDQSLGKPKCSE 3090  
Qy 339 APKFTGKRLISLDFQDVEIRTILOILAKESGMNIVASDSVNGKMTLSLKDVPWQALDLV 398  
Db 3091 EPKEPTGR-----DTPVGSWNVSPSQAPADIPTAI- 3120  
Qy 399 MQARNLDMEQGNINVIAPRDELLAKDAFLQAEKDIADLGLALYQNFQ----- 448  
Db 3121 -----PQNQNTSESPRTTSLKSTRTVQSSMPADDIDELABYDLOIARAVPTKHP 3171  
Qy 449 -----KYKNV-----BEPRSIIRLONADTTGNRNTLVSGRGSVL 482  
Db 3172 QPPPANQTPPQEPAPIDDRKXNIRPPLSEEBIIAFLINMDDDD-----AGNASGP 3222  
Qy 483 ID-----PATNTLIV-----TDTSRVIEKFKRLIDELDV-PAQOVMEARIV 523  
Db 3223 VDLHSVQAPKPLPKQSKPTTNQFVPLDWMWTETEPVVD-----ADSLDLPKQQLFSEW- 3275  
Qy 524 EAADGFSRD-LGVKFGATGCKKXKNDTSAFGNGV-----SGFGGDDKQWGAETKI 572  
Db 3276 -----STROLLNINVRDVRVYEEESDDEYTVSDQHLVPAVSPTSVSSYSDVTVSYTDI 3330  
Qy 573 NLPITA-----AANSI-----SLVRAI-----SSGALNLSAS 603  
Db 3331 NDRSVVCPDGNQNAQNVREFLDTHSSRVVVPADELLSSRRYFRSTLSAMALLAACRT 3390  
Qy 604 LSKTKTLANPRVLTO-NR 620  
Db 3391 IVR-RURATRRVLTDIR 3407

|            |  |   |                        |
|------------|--|---|------------------------|
| QY         | 294  | LNNDTQLIITTAGNWLNVKSAAPGYTTFQVLPKKQNLKESGGVNNAPKTTGRTKISLDFO  | 353                    |
| DB         | 36   | LHADTSL-----KSKNE---HYKSSDMVSKTDSIYIG--NSSFTYHGEPLPKLE        | 83                     |
| QY         | 354  | DVE---IRT-----ILOILAKESGMNIV---ASDSVNG-----KMTLSL             | 386                    |
| DB         | 84   | GVHGIILRSSTPLGFDVLSWIDSSGIPVKHTTKDVISGGVSKSLAATVAEKMSAT       | 143                    |
| QY         | 387  | KDVPWDQALDLMQARN-----LDMRQOGNI-----                           | 412                    |
| DB         | 144  | CGKSTDFDHLLEVSSEHQLMDVNYQALSTFLDKVAANYNLWYTESGRIAFSNEETK      | 203                    |
| QY         | 413  | ---VNIAPRODELLAKDAFLQAEKDIADLGALYSQN---FOLKYKVVEEPRSLRLDNA    | 465                    |
| DB         | 204  | RFSISILPGGYTKNSISSDSSSSSGSSGSSGSGAELKFDSDVDFWK-----           | 256                    |
| QY         | 466  | DTTGNRNTLVSGRGSVLIDPATNTLIIVTTRSVIEKFRKLIDELDVPAQ-QVMIEARIVE  | 524                    |
| DB         | 257  | DIENSIKILIGSDGYSISTSTSSVIVRTSSANKKINEVINTLNAQLERQVTDVAIYN     | 316                    |
| QY         | 525  | AADGFSRDLGVKFGATGKKLKNDSAFGVNWSFGGDDKMGAEKINLPITAAANSIS       | 584                    |
| DB         | 317  | VTTTSSDLAMSLEAL-----LKHNGGVLSVSTSNF-----AATSGTSPFTGYLNG--     | 363                    |
| QY         | 585  | LVRAISSGALNLELSAGESLSKTKTLANPRVLTONRKEAKIESGYEIPF-----TVTSIA  | 639                    |
| DB         | 364  | --NGDSSNQVLLNLAEK--GKVSVWTSASVTTMSGQPVLPKVGNDRTYVSEIGTVLSQS   | 419                    |
| QY         | 640  | NGGSTNTELKAVLGLTVTNITPDQI-----MTVKINKDSPACASGNQITLIS          | 693                    |
| DB         | 420  | STSTASTSTVTSGLFAMNLLQVADDDGNILQYGVTLSELVSGNSNGFDQATVNGTVIQLP  | 479                    |
| QY         | 694  | TKNLNT---QAWVENGGLIIVGGIYEE-----DNGNTLTAKVPLIGDIPVICNLFKTRGK  | 744                    |
| DB         | 480  | NVDSTTFVQSSMLRNGNTLVLAG-YEKENESVDQGVGTTTSFKLLG-----GAL-----NG | 529                    |
| QY         | 745  | KTDRELLIFITPRIM   | 760                    |
| DB         | 530  | SASRTVTVICITPRII  | 545                    |
| RESULT 33  |  |   |                        |
| BPFB_ECO27 | ID   | BPFB_ECO27  | STANDARD; PRT; 553 AA. |
| AC         | Q47068;  |   |                        |
| DT         | 28-FEB-2003  | (Rel. 41, Last sequence update)                               |                        |
| DT         | 28-FEB-2003  | (Rel. 41, Last annotation update)                             |                        |
| DE         | Outer membrane lipoprotein bfpB precursor (Bundle-forming pilus B).        |   |                        |
| GN         | BFPB.  |   |                        |
| OS         | Escherichia coli O127:H6.  |   |                        |
| OG         | Plasmid pMAR2.   |   |                        |
| OC         | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;          |   |                        |
| OC         | Enterobacteriaceae; Escherichia.   |   |                        |
| OX         | NCBI_TaxID=168807;   |   |                        |
| RN         | [1]  |   |                        |
| RP         | SEQUENCE FROM N.A.   |   |                        |
| RC         | STRAIN=O127:H6 / E2348/69;   |   |                        |
| RA         | MEDLINE=96310370; PubMed=8733231;  |   |                        |
| RX         | Stone K.D., Zhang H., Carlson L.K., Donnenberg M.S.;                       |   |                        |
| RT         | "A cluster of fourteen genes from enteropathogenic Escherichia coli is     |   |                        |
| RT         | sufficient for biogenesis of a type IV pilus."                             |   |                        |
| RL         | Mol. Microbiol. 20:325-337(1996).  |   |                        |
| CC         | -!- FUNCTION: Is absolutely required for pilus biogenesis, and for         |   |                        |
| CC         | EPEC localized adherence and autoaggregation. Acts at a step in            |   |                        |
| CC         | the BFP biogenic pathway after production and processing of the            |   |                        |
| CC         | structural pilus subunit bfpA (By similarity).                             |   |                        |
| CC         | -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid        |   |                        |
| CC         | anchor (By similarity).  |   |                        |
| CC         | -----  |   |                        |
| CC         | This SWISS-PROT entry is copyright. It is produced through a collaboration |   |                        |
| CC         | between the Swiss Institute of Bioinformatics and the EMBL outstation -    |   |                        |

Query Match 3.88; Score 145.5; DB 1; Length 552;  
Best Local Similarity 19.48; Pred. No. 0.36;  
Matches 108; Conservative 98; Mismatches 215; Indels 135; Gaps 23;

RESULT 32  
BPFB\_ECO11  
ID BPFB\_ECO11 STANDARD; PRT; 552 AA.  
AC Q8S142; Q46777;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Outer membrane lipoprotein bfpB precursor (Bundle-forming pilus B).  
GN BFPB.  
OS Escherichia coli O111:H-.  
OG Plasmid pB171.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=168927;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O111:H- / B171;  
RX MEDLINE=96196169; PubMed=8626330;  
RA Schel I., Puente J.L., Ramer S.W., Bieher D., Wu C.-Y.,  
RA Schoolnik G.K.;  
RT "Complete DNA sequence and structural analysis of the enteropathogenic  
RT Escherichia coli adherence factor plasmid."  
RL J. Bacteriol. 178:2613-2628(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O111:H- / B171;  
RX MEDLINE=99426847; PubMed=10496929;  
RA Tobe T., Hayashi T., Han C.-G., Schoolnik G.K., Ohtsubo E.,  
RA Sasakawa C.;  
RT "Complete DNA sequence and structural analysis of the enteropathogenic  
RT Escherichia coli adherence factor plasmid."  
RL Infect. Immun. 67:5455-5462(1999).  
RN [3]  
RP CHARACTERIZATION.  
RC STRAIN=O111:H- / B171;  
RX MEDLINE=97086623; PubMed=8932312;  
RA Ramer S.W., Bieher D., Schoolnik G.K.;  
RT "BfpB, an outer membrane lipoprotein required for the biogenesis of  
RT bundle-forming pili in enteropathogenic Escherichia coli."  
RL J. Bacteriol. 178:6555-6563(1996).  
CC -!- FUNCTION: Is absolutely required for pilus biogenesis, and for  
CC EPEC localized adherence and autoaggregation. Acts at a step in  
CC the BFP biogenic pathway after production and processing of the  
CC structural pilus subunit bfpA.  
CC -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
CC anchor.  
CC -!- INDUCTION: During exponential-phase growth; repressed by ammonium.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
CC -----  
CC EMBL; U27184; AAC4404.1; --  
CC EMBL; AB024946; BAA84840.1; --  
CC InterPro; IPR004846; GSPII\_11; Pilproteins.  
CC Pfam; PF00263; GSPII\_III; 1.  
CC PROSITE; PS00043; PROKAR\_LIPOPROTEIN; 1.  
KW Embria; Membrane; Outer membrane; Lipoprotein; Signal; Plasmid.  
FT SIGNAL 1 17 PROBABLE  
FT CHAIN 18 552 OUTER MEMBRANE LIPOPROTEIN BFPB.  
FT LIPID 18 18 N-ACYL DIGLYCERIDE (PROBABLE).  
FT DOMAIN 227 242 POLY-SER.  
FT CONFLICT 332 335 LLKH -> FYND (IN REF. 1).  
SQ SEQUENCE 552 AA; 58372 MW; F8CAE36171DCA956 CRC64;





Db 545 --TSTLSSPLGTSQVMTASGLQTAATAAALQAAQIPANASLAAMAAAGLNPGLMASS 602  
Qy 706 ---GGTLL-----VGGIYED--NGNLTVPKLL---GDIPVI-----GNL-FKTRGK 744  
Db 603 QFAAGALLSLNPGTLLGALSALMSNSTLATIQAALASSGUSPITSLDASGNLVPANAGG 662  
Qy 745 KTDRELLIFITPR 758  
Db 663 TPNIVTAPLENPQ 676

RESULT 36  
SCA4\_RICJA  
ID -SCA4\_RICJA STANDARD; PRT; 1018 AA.  
AC Q9AJ79; 006654;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)  
DE (Protein PS 120) (rps120).  
GN SCA4 OR D.  
OS Rickettsia japonica.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Rickettsiaceae; Rickettsieae; Rickettsia.  
OX NCBI\_TaxID=35790;  
RN Rickettsia japonica.  
RP SEQUENCE FROM N.A.  
RC STRAIN=YH;  
RX MEDLINE=20049841; PubMed=10585146;  
RA Uchiyama T.;  
RT "Sequence analysis of the gene encoding a spotted fever group-specific  
RT intracytoplasmic protein PS120 of Rickettsia japonica.",  
RL Microbiol. Immunol. 43:983-987(1999).  
RN [2]  
RP SEQUENCE OF 8-1012 FROM N.A.  
RA Sekeyova Z., Roux V., Raoult D.;  
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the  
RT gene D, coding for an intracytoplasmic protein.",  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBSJ databases.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AB003696; BAA20142.1; -;  
CC EMBL; AF185055; AAK30686.1; -;  
CC PIR; T30853; T30853.  
KW Antigen.  
FT CONFLICT 8 8 N -> D (IN REF. 2).  
FT CONFLICT 449 449 K -> E (IN REF. 2).  
FT CONFLICT 693 693 S -> G (IN REF. 2).  
SQ SEQUENCE 1018 AA; 111147 MW; F41F015392671BAA CRC64;  
Query Match 3.8%; Score 144.5; DB 1; Length 1018;  
Best Local Similarity 20.1%; Pred. No. 0.97;  
Matches 161; Conservative 101; Mismatches 321; Indels 217; Gaps 36;  
Qy 107 KPGQVNTVRGN-KVMIPINESDDTVSAPAPVAAAPAAKQGGCTVYQVRSIRIQ 164  
Db 183 KPVQWENHVSADLRATVVKNDAGDELCTLNNTTKTFFTLAKQDG--TQVQISSYR-E 239  
Qy 165 TLVPGKTTAA-APFTESVVSAP-FSPAKQAAASAKQQTAAKQQTAAKQQAAP 222  
Db 240 IDFPKLDAGSMHLSVALKADGTPKSKDKAVVFTAHYEGP-----NGKPOLKEISSP 295  
Qy 223 AKQTNIDFRKD-----GKNAGIIEAALGACQAPDISQOHDHIIVTL 264

Db 296 KPLPAGTGDAAIAYIEHGGIYTLAVTRGKYKEMKVELNQGSVDLSQAEIILIQG 355  
Qy 265 KNHTLP-----TTLQRSIDVADFK-----TPV-----QKVTLKLNNDTQ 299  
Db 356 QSKQPLITPQOTTSSVPEPPQYKQVPPITPTNOPLQPETSQMPQSQOVNPNLLNATA 415  
Qy 300 LIITTAGNWLWN-----KSAAPGYT---FQVLPKQNLESQGVNNAKTF 343  
Db 416 L---SGSMQDLNLYVYVAGLTKEDILKEAATAILNKKSDIAEKQANI-----IALAENTV 468  
Qy 344 TORKISLDFQVVEIRTIQILAKESGMNIVASDSVNGKMTLSL-KDVPWDQALDLVMOAR 402  
Db 469 NNKNLTPD-----AKVAGVNAVLETTIKNDQNTPDLEKSKMLEATVAIALNSE 515  
Qy 403 NLDMEQGNVNIAPRDELLAKDKAFLOAEKD-TADLGALYSQNFQLYK----- 451  
Db 516 NLEPKQKQIILEKAVDVGLSLKDDASRAAIDGITD--AVIKSNLSTEDKGTMTFVAGDK 573  
Qy 452 -NVEBFR-----SILR-----LDNADTTGNRNTLVSGRGSVL 482  
Db 574 VNVSELSNAEKQKLGSLVKKGVEAQLSPAQQQLAQOQLDKITAEQTKKDTIKKVNAIL 633  
Qy 483 IDPATNT-LIVTDRSVIEKFRKLIDELDVPAQOVNIEARIVEAADGFSRDLGVKFGATG 541  
Db 634 FDPPLSNTLKTNIQAIIIS-----NVLDGPA-TAEVKGEBIIQE-----ITNTVAG 677  
Qy 542 KKKLKNLDTSAFCGVNSG-----FGDDDKWGAETKINLP----- 575  
Db 678 SSLEAHDKAAIIGKISETIATHSDTSLPNKALIMASAEKGIASQANLPDRELMTKGL 737  
Qy 576 ---ITAAANSISLVRAISGA--LMELSASELSKTKTLANPVL---TQNRKEA---- 623  
Db 738 VDGIVEGKGPEITKAVSSGIDNSINDSEALKKAKDAASEAALDRETQNLTEGLKQG 797  
Qy 624 KIESGV---EIPFTVTSIANGSSNTTELKKAVLGLTVPNTITPDQOIIMTVKINKDSPA 680  
Db 798 NIEEHKPHDIIYNKREVINAVNPVIEALEK-----SKEPVVSAERIV-----QET 844  
Qy 681 QCASGNGTILCISTKNLNTQAMVENGTLIVGGIYEEDNGNTLVKVPILGDPVIGNLFK 740  
Db 845 SSLINNISKLAVEKVN-NFRMLSPNGNL---KTLKEKKEESIKVY-----DELVK 891  
Qy 741 TRGKKTDRRELLIFITPRIM 760  
Db 892 AFGTKSSTEEQSQSIKANLI 911

RESULT 37  
SCA4\_RICPR  
ID -SCA4\_RICPR STANDARD; PRT; 1022 AA.  
AC Q9ZD49; Q9AJ36; Q9ZD48;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)  
DE (Protein PS 120).  
GN SCA4 OR RP498/RP499.  
OS Rickettsia prowazekii.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Rickettsiaceae; Rickettsieae; Rickettsia.  
OX NCBI\_TaxID=782;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Madrid E;  
RX MEDLINE=99039499; PubMed=9823893;  
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
RA Scheritz-Ponten T., Alsmark U.C.M., Fodowski R.M., Naeslund A.K.,  
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
RT "The genome sequence of Rickettsia prowazekii and the origin of  
RT mitochondria".  
RL Nature 396:133-140(1998).  
RN [2]  
RP SEQUENCE OF 11-1016 FROM N.A.

RA Sekeyova Z., Roux V., Racult D.;  
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the  
RL 'gene B' coding for an intracytoplasmic protein.";  
CC submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
CC FRAMESHIFT IN POSITION 234.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AJ235272; CAAL4951.1; ALT FRAME.  
DR EMBL; AJ235272; CAAL4950.1; ALT FRAME.  
DR EMBL; AF200340; RAK31305.1; -  
KW Antigen; Complete proteome. EFDPL -> RGLV (IN REF. 2).  
FT CONFLICT 11 15  
FT CONFLICT 365 365 H -> Y (IN REF. 2).  
FT CONFLICT 413 413 MISSING (IN REF. 2).  
FT CONFLICT 957 957 G -> R (IN REF. 2).  
SQ SEQUENCE 1022 AA; 114410 MW; 03230E3A663A9622 CRC64;  
  
Query Match 3.8%; Score 144.5; DB 1; Length 1022;  
Best Local Similarity 19.8%; Pred. No. 0.97;  
Matches 171; Conservative 121; Mismatches 315; Indels 255; Gaps 39;  
  
QY 31 KVSLLPNKQXIVSPDKEL--VNPTGVTSSPARIALDFEQTGISMDOQVLEYAD---P 85  
DB 175 QVSASDLRSTVKNVDEGEELCTLNETHVTKD---LIVAKQDGTQV--QINSYREINFP 228  
  
QY 86 L-LSKISAQNSPARVLNLPKPGQ---YNT-----EVRGNKVMIFINE 126  
DB 229 IKLDKANGMHLMSWALKADGTPAKDKAVYFAHYEEGNGKFPQKEISSQPLKFVET 288  
  
QY 127 SDDTVS-----APARPAKVAAPAPAKQCGCRTVYQVRSIRI-QTLYPGKTTAA 174  
DB 289 GDDAVAYIEHGGEIYTLAVTRGVKXEMKEVALNHG-----QSVASGTTIAEDLTHVQ 341  
  
QY 175 APTESVSVSVAFPSPAKQAASAKQOTA-----APAKQOTAPAKQA--- 219  
DB 342 GPSHETHKPIIP-----NQELESSIEOHTSQVFPITTFNKSLQPKISQIHOLQPOAQS 397  
  
QY 220 -----AAPAKQTNID-----FRKD---GKNAGIIEALALGFAG--QPDISQ 256  
DB 398 SGTPNPVNLNANALSTSWQDLLNINSYLTKNODINKQSDLIKAEALIALNNKKSDFAEK 457  
  
QY 257 HDHIIVTLKNHPTTLQRLSDVADFKTPVQKVTLKLRLNDTQLIITAGNWLKNSAA 316  
DB 458 QYNIIDLAKN-----IFSNKDIIDAKVNVVNTLLETQNDQ----- 495  
  
QY 317 PGYFTFOVLPKQNLSEGGVNNAPKTTGRKISLDFQDVIRTILOILAKESGMNIVASD 376  
DB 496 -----TLDI--KSKILEDV-----AITLNSIELKQKQILEKVVIGLSIKD 539  
  
QY 377 SVNGKMTLSLKVDPWDQALDLVMOARLDMRQGNIVN-----IAPREL-----L 422  
DB 540 DIS-----RVVAVDSIMDTVIKS-----NIANEDKEKIFITVFQINSYFESNV 583  
  
QY 423 AKKAFQLAEQKIDADLGALYSOLFAYKXVEEPRSLRLDNADTTGNRNTLVSGRGSVL 482  
DB 584 AKQKLLDSILKKTAEQTVLSPEQQQLMNQN-----LDNITTEHTKDTTEKVNIL 634  
  
QY 483 IDPATNTLIVTDRSVIEKFKRLIDELDVPAQ-----QVMIEARIVEAADGFS 530  
DB 635 LEPLSNTALKTTNIQVMTS-----NVLDSVPQVIEKSKLIQVWTKTVAESALVEPKDKTE 689  
  
QY 531 RDLGVKFGATGKKL-KNDTSAFGNGVNSGGGDDKNGAETKINL-----PIT 577  
DB 690 IVKGI-----GKTIIVTSDTSLPLHDKVIMSGVAKGIVESKNLDBRELIAGLVGDIY 744

QY 578 AAANSILVPAISS--GALNLELSASELSKTKTLANPRVL-----TONRKEA 623  
DB 745 EAKGDNAVVAHAISSMIANSINQSEKALRKSQVSEKVLDKIEQLNLDRELKAQINES 804  
  
QY 624 KIBSGYEIPFTVTSIANGSGSINTELKAVLGUTVTPNITPDQJIMTVKINKDSPAQCA 683  
DB 805 KLHD--DIYNKTDQVANA-----LKNVITVLDNDSGQRG-----VSEAPKVS 847  
  
QY 584 S-----GNQILCISTXVNTQAVENGTLVGGIYEEDNGNTLTQVPLLDIPVIGNL 738  
DB 848 SLLNDISKRTIEKIN-----NLRAWLSQDGNL---KTFEEKDEATKKV-----DEL 891  
  
QY 739 FKTRGKKTDRRELLIFITPRIM 760  
DB 892 VKAFDNKSSTEEQNFIKSNLI 913  
  
RESULT 38  
OMPB RICRI  
ID OMPB RICRI STANDARD; PRT; 1654 AA.  
AC Q53047;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Outer membrane protein B precursor (168 kDa surface-layer protein)  
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOMPb)  
DE (rOMPb) [Contains: 120 kDa surface-exposed protein (Surface protein  
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].  
GN OMPB.  
OS Rickettsia rickettsii.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI TaxID=783;  
RN [1] SEQUENCE FROM N.A.  
RC STRAIN=R;  
RX MEDLINE=92167802; PubMed=1724278;  
RA Gilmore R.D. Jr., Cleplak W. Jr., Policastro P.F., Hackstadt T.;  
RT "The 120 kilodalton outer membrane protein (rOMP B) of Rickettsia  
RT rickettsii is encoded by an unusually long open reading frame:  
RT evidence for protein processing from a large precursor.";  
RL Mol. Microbiol. 5:2361-2370(1991).  
RN [2]  
RP SEQUENCE OF 279-1654 FROM N.A.  
RC STRAIN=R;  
RX MEDLINE=90136087; PubMed=2515418;  
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;  
RT "Cloning, expression and sequence analysis of the gene encoding the  
RT 120 kD surface-exposed protein of Rickettsia rickettsii.";  
RL Mol. Microbiol. 3:1579-1586(1989).  
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR  
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
CC LAYER WITH HEXAGONAL SYMMETRY.  
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X16353; CA34403.1; -  
DR PIR; S18227; S18227.  
DR InterPro; IPR006315; Autotransporter.  
DR InterPro; IPR005546; Autotransporter.  
DR Pfam; PF03797; Autotransporter; 1.  
DR TIGRPFAM; TIGR01414; autotrans\_bar1; 2.  
KW Antigen; S-layer; Cell wall.

FT CHAIN 1 1333 120 kDa SURFACE-EXPOSED PROTEIN.  
FT CHAIN 1334 1654 32 kDa BETA PEPTIDE.  
FT DOMAIN 1181 1188 POLY-THR.  
SQ SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;  
  
Query Match 3.7%; Score 144; DB 1; Length 1654;  
Best Local Similarity 20.6%; Pred. No. 2;  
Matches 179; Conservative 105; Mismatches 323; Indels 262; Gaps 41;  
  
QY 43 KVSFDEINVTGFTSSPARIALDFEQT-----GISMDQOQVLEY 82  
DB 4 KPNFLKKLIS-AGLVITASTATIVASFAGSAMGAIIQONRTTNGAATTVDGAGFDQT-----58  
  
QY 83 ADPELLSKISAQN---SSRLVNLNKP-GQYNTVEVRGNKWKWIFINESDDTVSAPARA 138  
DB 59 AAP--ANVGVALNAVITANANGINFNPAGSFNGLLNTANNLAVTVSEDTTLGI---113  
  
QY 139 VKAAPAAPAQOQCRTYQVRSRIQTLPGKTTAAAPFTESVVSAPSPSPAKQQAAS 198  
DB 114 -----TNVYVHNSFNL-TLNAGKTLT---ITGQGV-----NAQAAT 148  
  
QY 199 AKQOTAPAKQOTAPAKQAQAPAKQNTNDFRKGKGNAGIIEALGAFAGQDPSQOHD 258  
DB 149 KNAQNV-----VQF-----NNGAIDNNDLKGVRIDFGAPAS 182  
  
QY 259 HIIVTLKNHTLPTTLQSLVDADPKTPQVKV--TLKELNNDTQIIITAGNWLNVKSA 316  
DB 183 TLVNLN--PTTKAPLILGDNAVANGVGLNVTNGFIQVSNKSPATVKAINIADG 239  
  
QY 317 PG--YFTFQVLPKQNIUESG--VNNAPKFTTGRKISLDFQDVETIILQILAKESGNI 372  
DB 240 QGIIENTDANNANTLNLQAGGTTINFTGDTGRLV-----LLSKHAAATNFI 288  
  
QY 373 VASDSVNGKMTLSKDVFPWD-----QALDLVNQARNLDMRQGNIVN-----IAPRD- 419  
DB 289 TGSLOGNLKGVIENTVAVDQLTANAGANAVITGNAGRAAGFVVDNGKVAITDG 348  
  
QY 420 ELLAKDAFLQAE-----KDIADLGLYVQNLQKYNV-----EPRSI---459  
DB 349 QVYAKDMVIOQANATQGVNFHFHVDVGADGTAFKTAASKVTITQDSNFGNTDFGNLAAQ 408  
  
QY 460 LRLDNADT-TGNRNTLVSGRS---VLIDPATNTL-----IVTDTSRVIEKFKLI 506  
DB 409 IKVPNAITLNGFTGDSNPNTAGVITFDANGTLESASADANVAVTNNTAIEASGAV 468  
  
QY 507 DELDVP-----AQQVMEARIVEAA-----DG-----FSRDLG 534  
DB 469 VQLSGTHAAELRLGNAGSIFKLADGTVINGKVNQTLVGGALAAAGTITLDGSATITGDIG 528  
  
QY 535 VKFGATGKK--LKNQ---TSAFGWVNSFGGDDKWAETKINLPITAAANSISLVRAI 589  
DB 529 NAGGAALQRTILANDAKTLTLGAGNIIGAGG-----TIDQANGTILKLTSTQ 579  
  
QY 590 SSGALNLISAS---ESLSKTKTLANPRVLTONKREAKIESG-----YEIPFTVTSIAN 640  
DB 580 NNIVVDFDLATADQGVVDASSLTNAQTTLNGKIGTIGANNKTLGQFNITGSSKTVLSN 639  
  
QY 641 GGSSTNTELKAVLG-----LVTPEITPDQIIMTVKINKDSP-----679  
DB 640 G-----NVAINELVINGDGAQFAHDTYILTRTNAGQCKIFPVPVNGTTLAAGTNLG 695  
  
QY 680 -----AQCASGNQ-----TIL-----CISTKRLNTQAMVEN-GGTLIVGGIYE 716  
DB 696 SATNPLAEINFGSGVNVVDVTLNVGEGVNLVATNITTTDANVGSFVNAGTNIVSGTVG 755  
  
QY 717 ENGNL-----TKVPLLDIPVGN 737  
DB 756 GQGGKNTVALENGTIVKFLGNATFNGN 784  
  
RESULT 39  
SED4\_YEAST STANDARD; PRT; 1065 AA.  
ID SED4\_YEAST

P25365;  
01-MAY-1992 (Rel. 22, Created)  
01-MAY-1992 (Rel. 22, Last sequence update)  
28-FEB-2003 (Rel. 41, Last annotation update)  
SED4 protein.  
SED4 OR YCR067C OR YCR67C OR YCR901.  
Saccharomyces cerevisiae (Baker's yeast).  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
NCBI\_taxid=4932;  
[1]  
SEQUENCE OF 1-514 FROM N.A.  
Contreras R., Demolder J., Fiers W., Molemans F.;  
Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
[2]  
SEQUENCE OF 446-1065 FROM N.A.  
MEDLINE=92221691; PubMed=1561837;  
Benit P., Chanet R., Fabre F., Faye G., Fukuhara H., Sor F.;  
"Sequence of the sup61-RAD18 region on chromosome III of  
Saccharomyces cerevisiae.";  
Yeast 8:147-153(1992).  
[3]  
SEQUENCE FROM N.A. AND CHARACTERIZATION.  
MEDLINE=93011014; PubMed=1327759;  
Hardwick K.G., Boothroyd J.C., Rudner A.D., Pelham H.R.B.;  
"Genes that allow yeast cells to grow in the absence of the HDEL  
receptor.";  
EMBO J. 11:4187-4195(1992).  
CC -!- FUNCTION: WHEN PRESENT IN MULTIPLE COPIES, ALLOWS CELLS TO GROW  
IN THE ABSENCE OF HDEL RECEPTOR BY EITHER SLOWING TRANSPORT FROM  
THE ER OR BY STIMULATING VESICLE BUDDING FROM GOLGI MEMBRANES.  
CC -!- SUBCELLULAR LOCATION: Type II membrane protein.  
CC -!- MISCELLANEOUS: IN THE PROCESS OF TRANSPORT, SED4 ITSELF MAY  
MIGRATE TO THE GOLGI APPARATUS AND FUNCTION IN SUBSEQUENT  
TRANSPORT EVENTS.  
CC -!- SIMILARITY: IN THE 2 WD REPEATS.  
CC -!- SIMILARITY: IN THE N-TERMINAL HALF, TO YEAST SEC12.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X59720; CAA42273.1; -;  
PIR; S19482; S19482.  
DR SED4; S0000653; SED4.  
DR GO; GO:0006888; P:ER to Golgi transport; IGI.  
DR InterPro; IPR000886; ER target.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00400; WD40; 2.  
DR SMART; SM00320; WD40; 2.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; FALSE\_NEG.  
DR PROSITE; PS00082; WD\_REPEATS\_2; FALSE\_NEG.  
DR PROSITE; PS0294; WD\_REPEATS\_REGION; 1.  
KW Transport; Protein transport; Glycoprotein; Endoplasmic reticulum;  
KW Transmembrane; Signal-anchor; Golgi stack; Repeat; WD repeat.  
FT DOMAIN 1 346 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 347 365 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
FT DOMAIN 366 1065 LUMENAL (POTENTIAL).  
FT REPEAT 259 298 WD 1.  
FT REPEAT 302 341 WD 2.  
FT DOMAIN 467 476 POLY-SER.  
FT DOMAIN 579 590 POLY-SER.  
FT DOMAIN 824 863 4 X 10 AA TANDEM REPEATS.  
FT REPEAT 824 833 1.  
FT REPEAT 834 843 2.  
FT REPEAT 844 853 3.  
FT REPEAT 854 863 4.  
FT CARBOHYD 388 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT



FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1039 1039 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT SITE 1062 1065 PREVENT SECRETION FROM ER (POTENTIAL).  
SQ SEQUENCE 1065 AA; 114079 MW; 071DA80D428065A8 CRC64;  
  
Query Match  
Best Local Similarity 18.4%; Pred. No. 1.3;  
Matches 144; Conservative 142; Mismatches 303; Indels 193; Gaps 35;  
  
QY 3 TLTITLISGLFVAT-----AAQTASAGNITDIKVSLLPNKQIKV--SPDKIVN-- 52  
DB 409 TESADIISATDVASDIETESFSDTSMRTTE-----DEQFVWISSADSQTSAD 461  
QY 53 -PTGFTVTSPIALDFEQTGISMDOOVLEYADPELLSKISAAGNSRRARLVMLNPKGV 111  
DB 462 IPTSASSSSSSSSSFVEEVTN-----EPIVS-----SPTSEITKELASPTPE 505  
QY 112 NTEVRGNKWIFIN-ESDDTVSAPAPAVKAAAPAAKQOCCTVTVQVRSIRIQTLYPGK 170  
DB 506 NIV---EKSLPLNSESIDLSSSNSITPYEPTPDLES-----KLSGLVIEQSESEI 556  
QY 171 TTAAPPTESVWSVAPSPFPAKQOAAASAKQOATAAPAKQOATAAPAKQOAAAPAKQT--- 226  
DB 557 TTDRESVSKLLSTESPLSHMPSSSSSSLSLSSLTSPPTALSTSTATAVTTQTNPIN 616  
QY 227 ---NIDPRKDGKNGAII-----LAALGFAGQP---DISQOHDHIIIVLKNHTL-P 270  
DB 617 DAANTSFLDNKSPASTREIYKTKIITEVITKIEYRNPASDSNAEAEQVYVTSSSMLLTP 676  
QY 271 TTLQSLDVAADFTPVQKVLTKLINDTQLIITAGNWLNVKNSAAPGVFTFOVLPPKKQ 330  
DB 677 TDMVSPVSEI-DPTAS-ELERVETPTHSISAEFDSVSNLPE-----EILSTSAS 731  
QY 331 LESGVNNAKPTTGRKISLDFQVEIRTL-OILAKESGMNIVASGVNGKWTLSKDV 389  
DB 732 QDS--ISSHPSTFSDSITSGFOSIEVSTVTSSVLASESIPSI--SDSTFSKF----- 780  
QY 390 PWDQALDVMQARNLDMRQGNVNIAPRDELLAKDKFLQAEKDAIDGALYSQNFOLK 449  
DB 781 -----HSIEPVSALVETA-----TSSSEKTKTKSRVIAFSTEDSERS 820  
QY 450 YKNVE---EPRSILRLDNADTNGRNTLVSGRGSVLIDPATNTLIVTDSRVIEKPKLID 507  
DB 821 SALINDSEYTSVL-ADNLEPT-----SVLADNSEPTSVLADSEPTSVF--TD 865  
QY 508 ELDPVQAQVMEARIVEAAGDFSDLVGKFGATGKKLKNDTSAFGMGVNSGFGGDKWG 567  
DB 866 AVQSPKTSV-----GQSSLSESTNIEGTSMA5----- 892  
QY 568 AETKINLPITAAANSISLVRAISSGAINLSELSK-----TKT-----LANPRVLQ 618  
DB 893 -----MIFSSSGASICALSDICKGLTSVE-SASSTVAQPMQGVTTTAPSFVSSPHKISA 945  
QY 619 NRKEAKIESY---EIPFTVTSIANG-----GSSTNTELKXAVL-----GLTV 658  
DB 946 SSIDA---SGFVQKEIMEIVQSSKDSSEAFGVBRHKISENVNTPVSRMLTTEMQASGVTDV 1002  
QY 659 TPNTIPDQIMTVKINKDSPACASQNTILCISTKNLNTQAVN-----GGTLIVGGIV 715  
DB 1003 TEDVSLSEVISALNVEITSLPNFVAPPQTIAAPLNNSTNIVNDNAVAGTVNAGLH 1062  
QY 716 EE 717  
DB 1063 DE 1064

RESULT 40  
MAPX\_DROME  
ID MAPX\_DROME STANDARD; PRT: 1185 AA.  
AC P23226; Q9V9S1;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE 205 kDa microtubule-associated protein.  
GN MAP205 OR CG1483  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91115949; PubMed=1703540;  
RT Irminger-Finger I., Laymon R.A., Goldstein L.S.B.;  
RT "Analysis of the primary sequence and microtubule-binding region of  
RT the Drosophila 205K MAP.";  
RL J. Cell Biol. 111:2563-2572(1990).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM B3).  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Ceincker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Fabros B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertler S., Fleischmann W.,  
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hoslin D., Houstoun K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusekern D.R., Paclet J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE REGULATION OF  
CC MICROTUBULE ASSEMBLY AND INTERACTION  
CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH CYTOPLASMIC MICROTUBULES AND  
CC WITH THE MITOTIC SPINDLE.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Name=J5;  
CC IsoId=P23226-1; Sequence=Displayed;  
CC Name=C2;  
CC IsoId=P23226-2; Sequence=VSP\_004319, VSP\_004320, VSP\_004321;  
CC Name=B3;  
CC IsoId=P23226-3; Sequence=VSP\_004319;  
CC -!- MISCELLANEOUS: PHOSPHORYLATION OF VARIOUS SERINE RESIDUES MAY PLAY  
CC A REGULATORY ROLE. THE BASIC DOMAIN CONTAINS NUMEROUS SEQUENCES  
CC THAT MATCH KNOWN CONSENSUS SEQUENCES OF SEVERAL DIFFERENT PROTEIN  
CC KINASES.  
CC -----





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 10:24:48 ; Search time 21 seconds  
(without alignments)  
1549.383 Million cell updates/sec

Title: US-09-701-271A-2

Perfect score: 3848

Sequence: 1 MNTKLTIIISGLFVATAAFQ.....ELLIFITPRINGTAGSLRY 769

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/protdata/1/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/protdata/1/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/protdata/1/iaa/6A\_COMB.pep.\*

4: /cgn2\_6/protdata/1/iaa/6B\_COMB.pep.\*

5: /cgn2\_6/protdata/1/iaa/PCOTUS\_COMB.pep.\*

6: /cgn2\_6/protdata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description            |
|------------|-------|-------------|--------|-------|------------------------|
| 1          | 1000  | 26.0        | 739    | 4     | US-09-328-352-6048, Ap |
| 2          | 980.5 | 25.5        | 751    | 4     | US-09-252-991A-33073   |
| 3          | 356   | 9.3         | 753    | 4     | US-09-252-991A-28934   |
| 4          | 341.5 | 8.9         | 649    | 3     | US-08-911-853-15       |
| 5          | 341.5 | 8.9         | 649    | 3     | US-09-479-409-15       |
| 6          | 341.5 | 8.9         | 649    | 4     | US-09-479-453-15       |
| 7          | 301   | 7.8         | 761    | 4     | US-09-328-352-5650     |
| 8          | 300   | 7.8         | 828    | 4     | US-09-252-991A-30225   |
| 9          | 264   | 6.9         | 995    | 4     | US-09-252-991A-22297   |
| 10         | 252.5 | 6.6         | 752    | 4     | US-09-328-352-5723     |
| 11         | 233   | 6.1         | 561    | 4     | US-09-198-452A-744     |
| 12         | 212.5 | 5.5         | 754    | 4     | US-09-198-452A-874     |
| 13         | 182.5 | 5.0         | 594    | 4     | US-09-252-991A-26461   |
| 14         | 183.5 | 4.9         | 1246   | 4     | US-09-252-991A-23140   |
| 15         | 160.5 | 4.2         | 1073   | 4     | US-09-206-942-49       |
| 16         | 160.5 | 4.2         | 1079   | 4     | US-09-206-942-47       |
| 17         | 157.5 | 4.1         | 1089   | 4     | US-09-328-352-5723     |
| 18         | 155.5 | 4.0         | 1095   | 4     | US-09-206-942-69       |
| 19         | 155.5 | 4.0         | 1536   | 1     | US-08-038-682-2        |
| 20         | 155.5 | 4.0         | 1536   | 1     | US-08-302-832-2        |
| 21         | 155.5 | 4.0         | 1536   | 2     | US-08-530-198-2        |
| 22         | 155.5 | 4.0         | 1536   | 2     | US-08-469-880-2        |
| 23         | 155.5 | 4.0         | 1536   | 2     | US-08-728-470-2        |
| 24         | 155.5 | 4.0         | 1536   | 2     | US-08-617-697-2        |
| 25         | 155.5 | 4.0         | 1536   | 3     | US-08-719-641-2        |
| 26         | 155.5 | 4.0         | 1536   | 4     | US-09-206-942-67       |
| 27         | 152.5 | 4.0         | 990    | 4     | US-09-252-991A-32469   |

|    |       |     |      |   |                    |                    |
|----|-------|-----|------|---|--------------------|--------------------|
| 28 | 150   | 3.9 | 1338 | 2 | US-08-728-470-9    | Sequence 9, Appli  |
| 29 | 150   | 3.9 | 1338 | 3 | US-08-719-641-9    | Sequence 9, Appli  |
| 30 | 150   | 3.9 | 1599 | 2 | US-08-617-697-9    | Sequence 9, Appli  |
| 31 | 147.5 | 3.8 | 1838 | 3 | US-09-120-863-2    | Sequence 2, Appli  |
| 32 | 146.5 | 3.8 | 734  | 4 | US-09-328-352-4412 | Sequence 65, Appli |
| 33 | 146.5 | 3.8 | 1180 | 4 | US-09-206-942-65   | Sequence 28, Appli |
| 34 | 146.5 | 3.8 | 1188 | 4 | US-09-206-942-63   | Sequence 63, Appli |
| 35 | 143.5 | 3.7 | 1220 | 4 | US-09-206-942-28   | Sequence 26, Appli |
| 36 | 143.5 | 3.7 | 1226 | 4 | US-03-206-942-26   | Sequence 10, Appli |
| 37 | 143   | 3.7 | 1600 | 2 | US-08-617-697-10   | Sequence 35, Appli |
| 38 | 142.5 | 3.7 | 915  | 4 | US-09-206-942-35   | Sequence 37, Appli |
| 39 | 141.5 | 3.7 | 1222 | 4 | US-09-206-942-37   | Sequence 34, Appli |
| 40 | 141.5 | 3.7 | 1228 | 4 | US-09-206-942-34   | Sequence 4, Appli  |
| 41 | 141   | 3.7 | 1912 | 1 | US-08-409-895-4    | Sequence 4, Appli  |
| 42 | 141   | 3.7 | 1912 | 3 | US-08-685-467-4    | Sequence 4, Appli  |
| 43 | 141   | 3.7 | 2353 | 3 | US-09-377-155-33   | Sequence 33, Appli |
| 44 | 141   | 3.7 | 2353 | 3 | US-08-913-942-4    | Sequence 4, Appli  |
| 45 | 141   | 3.7 | 2353 | 4 | US-09-669-974-33   | Sequence 33, Appli |

ALIGNMENTS

RESULT 1  
US-09-328-352-6048  
; Sequence 6048, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03FA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 6048  
; LENGTH: 739  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-6048

|                       |                   |   |             |             |
|-----------------------|-------------------|---|-------------|-------------|
| Query Match           | 26.0%             | Score 1000;   | DB 4;       | Length 739; |
| Best Local Similarity | 30.5%             | Pred. No. 2.4e-72;  |             |             |
| Matches 242;          | Conservative 146; | Mismatches 266;   | Indels 140; | Gaps 12;    |
| QY                    | 14                | VATAFQTASAG-NITDIKVSLLPNKQIVKVSFDEKIVNPTGVTSSPAIALDFETG     | 72          |             |
| Db                    | 31                | VAIATMQAASQVSMNTIVPMQIAGQGTETIRVMFNGLPFPQAYOLENPSRLILDFDQAK | 90          |             |
| QY                    | 73                | ISMDOQVLEYADELLSKISAQNSRARLVNLNKPQGYNTEVRGNKVMIFINESDDTVS   | 132         |             |
| Db                    | 91                | QQLKQSKISVATNEASSVDVTSDDORSRLTVNLKDAGAFTRVEGNTFILKINS-----  | 144         |             |
| QY                    | 133               | APARPAVKAAPAAKQGGQRTVYQVRSIRIQTLYPGKTTAAAPFTESVSVSAPFSPAK   | 192         |             |
| Db                    | 145               | --AQTSNKPLPVVSAQPGV-----                                    | 162         |             |
| QY                    | 193               | QQAASAAQQTAAAPAKQQAAPAKQQTNIQDKGQKAGITELAALGFAGQPD          | 252         |             |
| Db                    | 163               | -----SNIGFQSGSGEGVVVLLGSNTVPD                               | 189         |             |
| QY                    | 253               | ISQHQDHIIVTLKNHTLPTTLQSLDVAFKTPQVKVTKLRLNNDTQIITAGWELVN     | 312         |             |
| Db                    | 190               | VQOQSGKVIRTIGTKIPTHLARRLVNDFATPVSSIDAYNDKGVGVTIQSSGSYE---   | 246         |             |
| QY                    | 313               | KSAAGYTFQV-----LPKKQLSEGGVNNAPKTTGPKISLDFODVEIRTIQLA        | 365         |             |
| Db                    | 247               | -----YWAYQENKLTISLKRPODKNVTSLYKTP-NYSGNKLSDLPQDIEVRRVQLQA   | 299         |             |
| QY                    | 366               | KESGMNIVASVNGKMTLSLKDVPWDQALDLVQARNLDRQCGNIYNIAPRDELAKD     | 425         |             |
| Db                    | 300               | DFTGIMVAADSVQGNITLRLKQVDPWDQALDILKTKULDKRRNGNVIAPVAELIKAE   | 359         |             |

```

426 KAFLOAEKDIA DLGALYSQNPQLKXKVVEFRSILRLDNADTTG-----NRN 472
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
360 EEEAKAVASQVKLAPLOTEIQLKYAKAGIMGLITQGSNNNSGLHOTSGGTTSTNLN 419
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
473 T-----LYSGRGSVLIPATNTLAVTSTRSVEIEFKRLIDELDVPAQQVWIEAR 521
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
420 TGVDLSGNVWGLSLPSRGITIQDDORTNLTINDTAQSIDQIRKMDLLDVQVQVWIEAR 479
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
522 IVEAADGFRSRLGVKFKATGKKLKNDTSAFG-----MGVNSGFGGDDKWAETKINLP 575
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
480 IVRASTSTFKELGVKXWGLSGITNNNNLLVGGSEITLWNLREPKKDEITGGYKYTIERP 539
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
576 -----ITAAANSISL-VRAISSGALNIELSASELSKTKTLANPRVLTVQNRKEAKIE 626
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
540 DNLNVDLGVSNPAGSIAPGLISMSDFMLDLSEALQADGYGEIVSTPKVNTADKOTAKVA 599
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
627 SCYEIPFTVTSIANGCSSTNTLKKAVLGLITVTPNITPDGQIIMTVKINKDSPACASGN 686
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
600 TCGQVPYOSTNSAAGSTATTSFKDALLSNLVTPSTIPDGKIOMKLDISKDSVAGAAPNG 659
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
687 QPILCISTKNLTQAMVENGCTLIUVGGIYEBEDNGNTLTUKVPLIGDIPVTGNLFKTRGKKT 746
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
660 EUIL--NKNNTNVLVNGETVILGVFFGQTTMNSQTKVPFFGDIPIVVGRULFRKDVKSD 717
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
747 DRRELLIFITPRIM 760
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
718 DXOELLIFVTRIV 731
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 2  
US 9,252,991A-33073  
; Sequence 33073 Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION: Rubenfield et al.  
; APPLICANT: Marc J.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

; FILE REFERENCE: 107/198-136
; CURRENT APPLICATION NUMBER: US/09/252,991A
;
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
;
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
;
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
;
; SEQ ID NO 33073
;
; LENGTH: 751
;
; TYPE: PRT
;
; ORGANISM: Pseudomonas aeruginosa
;
; US-09-252-991A-33073

```

|                       |                   |                    |             |             |
|-----------------------|-------------------|--------------------|-------------|-------------|
| Query Match           | 25.5%             | Score 980.5;       | DB 4;       | Length 751; |
| Best Local Similarity | 30.7%             | Pred. No. 9.2e-71; |             |             |
| Matches 245;          | Conservative 152; | Mismatches 270;    | Indels 131; | Gaps 20;    |

```

QY 1 MNTKLTIIISLFLVATRAFTA--SAGNITDIKVSSSLPNKQKIVKVSFDPKKEIWNTPGFVTS 59
Db 38 MNSGLSKL--GIALLAAMFALLAAALEKLDVAALPGDVRVELKQFDEPVAAPRGTYIE 95
QY 60 SPARIALDFEOTGISMDOQVLEYADPULSKISAAQNSRARLVNLNKNKPGOYNTFEVRGNK 119
Db 96 QPARIALDLPGVQNKLTGKRELSVGNTRSVTVVEAKDRTLIINLTSLSSYTTVRGENN 155
QY 120 VWIIFNSSDDIVSAPAPFAVKAAAPAAKQOQGRITVQVRSIRIQTLTPGKTTAAAPTE 179
Db 156 LFVVVGNH-----PAGASV--ASTAPVK----- 176
QY 180 SVVSVSAPFSPAQOAAASAKQOQTAAPAKOQTAAPAKQOAAAPAKQT--NIDFRKDGKNA 237
Db 177 -----ASPAPSAQPIKPKYVPAGAIRINIDFORGEKGE 212
QY 238 GIIELAALGFAGQPDISOQHDIHIVTLKNHTLPTLQRSLOVAFKTPYQVKVTLKRLNND 297

```

```

213 GNWVLDSDPTLSIQEOGGKIRLDFAKTQLPDALRVLDVKDFATPVQVFNAQAQSDR 272
298 TQLIITAGWE- LYNKSAAPGYFTFOVLP-----KKONLESGGVNNAPKFTGCRK 347
273 TSITIEPSGLYDLYVT--DNRLTVSKPMTTEDAERKKDNF-----AYTGK 320
348 ISLDFODVIRITLQILAKESGMNIVASDVNGKMTLSLKDVDPQALDVLNQARNLDMR 407
321 LSLNFODIDVRSLVQLIADFTDLNLVASDVTQGNITLRLQNVDPQALDVLKTKGLDKR 380
408 QQGNIVNIAPRELLIAKXAFQAKSDIADLQALYSONPOLKYKXWBEFRSLRLDNADT 467
381 KLGNVLLVAPADEIAREROELEAQOIAELAPLRRELIQVNYAKAADIAKUFSQVTS- 439
468 TGNRNLTVSGRGSVLDPATNLTIVTDRSVIEKFKRLDELDPVPAQQVMIEARIVEAAD 527
440 -GGQEKGEGRGSITVDDRTNSIIAYQPERLDELRRIVSOLDIPRVQWIEARIVEANV 498
528 GFSRDLGVGFAGTKKKLKNDTSAFGWGVNSFGGDDKKGAAET-----KINLPITAAANS 582
499 GYDKSLGVWGGAYHXGNWN----GYGDKGNIGHKDEDMGCGPIAGNCTPFTGTSTKS 553
583 ISL-----VRATSSG-----ALNELSASLSLSTKTLANPRVLTQNRKKAIES 627
554 PSFPVLDGAKDATSGIGIGIFTDNIILDLQLSAMEKTNGEIVSQPKVTSDEKATAKILK 613
628 GYBIPPTVTSIANGSSSTTELUKXAVLGTVTPTNITPDQCIIMTVKINKDSQAQCSGNQ 687
614 GSEVPY--CEASSSGATSTSFKEAALSLEVTPOITPDNRIIIVEKVTKDAP-----DYQ 665
688 TIL----C-STKNLNTQAMVENGTLIVGGIVEEDNGNLTITKVPPLGDPVIGNLFPKTRG 743
666 NMLUNGVPPIKINEVNAKILVNDGETIVIGVFSNEQSKSVEKVPFFGELPYLGRLEFR-RD 724
744 KKTDRR-ELLIFITPRIM 760
725 TVTDRKNELLVFLTPRIM 742

```

```

RESULT 3
US09252991A-28934
; Sequence 28934 Application US/09252991A
; Patent No 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

```

; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; CURRENT APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 3312
; SEQ ID NO 28934
; LENGTH: 753
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-28934

```

|                       |       |  |                 |             |
|-----------------------|-------|--|-----------------|-------------|
| Query Match           | 9.38; | Score 356;   | DB 4;           | Length 753; |
| Best Local Similarity | 23.1; | Pred. No. 3.3e-20;   |                 |             |
| Matches               | 174;  | Conservative 129;  | Mismatches 257; | Indels 192; |
| Gaps                  | 33;   |  |                 |             |
| QV                    | 135   | ARPAVKAAPA--APAKOQGGRTVYQVR-----SIRIQTLY-----                  | 167             |             |
| Db                    | 14    | AAGAWKACPSRTVPGDCWPARTTQATRTSRNNCKACRMRTLQHCASAWTFCASRWKPRP    | 73              |             |
| QV                    | 168   | ---PGKTTAAAPFTESVYVSVAAPSQAQOQAAASAKQOTAAPAKQOATAAPAKQOAAAPAK  | 224             |             |
| Db                    | 74    | SONEPKKTTRASP--RPRTKATDTPNSOPIRALFAPSSRSYVPAVLISLALGICIAAHA--- | 129             |             |





QY 493 -----TDRSVIE----- 500  
Db 250 AAFORLANARSLDIPSTRANARVIRLRHSDAKSLAETLGDISEGLKTAEGGEAASSK 309  
QY 501 -----KPKLIDELVPAQOVMEIARIVEAADFSRDLGV 535  
Db 310 PQNILRADESALVALADPDVATLEEIVRLNDVPAQVMEAAIVEISGDISALGV 369  
QY 536 KFGATGKKLKNDTSFAFGWVNSGFGG-----DDKWAETKINLPITAA 579  
Db 370 QWAVDA---RGGTGGUG-GVNFNTGLSVGTVLKAIONEEIPDD-----LTLPL 413  
QY 580 ANSISLVRAISSGALNLELSASLSKTKTLANPRVLTQNEKEAKIESGVEIPEF---TWT 636  
Db 414 DGAIGIGTENFGALITALSAN---SASNLSLTPSLTLDNQAEILVGNQVFPFGSYT 470  
QY 637 SIANGGSSTTEKKAVGLGT--VTPNITPDGQIIMTV--KINKSPAOCASGNQIILCI 692  
Db 471 TDASGANNPTTIEREDIGVTLKVTPHINDGATLRLEVEQEISSIAPSAGWAQAVDLVT 530  
QY 693 STKULNTQAMVENGSTLIVGVEEDNGNTLTKVPLLDGDPVIGNLTKRGKTKDRRELL 752  
Db 531 NKRSIKSVIADGGQVILVGLIQDDVTSDTSKVPLLDGDPVIGLGRFSTRKTHVKRNLN 590  
QY 753 IFITPRIM 760  
Db 591 VFLRPTIV 598  
RESULT 7  
US-09-328-352-5650  
; Sequence 5650, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; PRIOR FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 5650  
; LENGTH: 761  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-5650  
Query Match 7.8%; Score 301; DB 4; Length 761;  
Best Local Similarity 23.08; Pred. No. 9, 5e-16;  
Matches 114; Conservative 90; Mismatches 186; Indels 105; Gaps 18;  
QY 331 LESGVNNAKPTTGRKISL-----DFQD-VEIRTILOILAKESGMNIVASD 376  
Db 214 LEAMATGASKDFSGARIRIADNRNTRILIKGDPQTRKRIRHMIEML-----DVPSAD 267  
QY 377 SVNGKMTLSLKDVFDQALDVMQARNLDMRQOGNIVNIAPRDELLAKDKAFLOAEKDIA 436  
Db 268 RLGLKVFRIK-----YASAKNLSEILQGLVTG-----QAVSSNNNSNNSSNS 312  
QY 437 DLGALYSQNFQKYNVEFRS-----ILRLNADTNGNRLTVS---GRGSVLIDPAT 487  
Db 313 PINSLIGNN-QNSGNTSGSGTSISPTAINLNGNSNNQNNITSFNONGVSIADNAQ 371  
QY 488 NTLIVTDRSVIEKRLIDELVPAQOVMEIARIVEAADFSRDLGVKFGATGKKLKN 547  
Db 372 NSLVVKADPOLMEIESAIQDLVRQOVLEIAAIEVSKDKADQGVOM----- 421  
QY 548 DTAFNGVNSGFGDDKWAETKINLPITAAANSISLVRA---ISSGALNL--ELSASBS 603  
Db 422 ---ALG-DINSIG-----LIN--FTNAGSSLASLAAGVLTGGAAGLGAIGAGSS 466  
QY 604 LSXTK-----TLANPRVLTQNEKEAKIESGVEIPEFVTS 637

Db 467 IALGKYKEGADGSRQLYGALIQALKENTASNLLSTPSTVMTDNEBAYIVVGONVPEFTGS 526  
QY 638 IANGSSSTN--TELKAVGLGT--VTPNITPDGQIIMTVKINKDSPAOCASGNQIILCIS 693  
Db 527 VTTNSTGINPYTTVERKQDVGLKVIPIHIGENTVRLIE-CEVSNVQASKQAAADLTN 585  
QY 694 TKNLNTQAMVENGSTLIVGVEEDNGNTLTKVPLLDGDPVIGNLTKRGKTKDRRELLI 753  
Db 586 KRAIKTAVLAERHGTWILGSLVSDDEFEFNRQOIGPLSSIPYLGRFLFRSDTRSNKRLNV 645  
QY 754 FITPRIMGTAGNSLR 768  
Db 646 FIHPTIVGDANDVRR 660  
RESULT 8  
US-09-252-991A-30225  
; Sequence 30225, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30225  
; LENGTH: 828  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30225  
Query Match 7.8%; Score 300; DB 4; Length 828;  
Best Local Similarity 21.4%; Pred. No. 1, 3e-15;  
Matches 122; Conservative 97; Mismatches 180; Indels 170; Gaps 16;  
QY 349 SLDFQDVEIRTILOILAKESGMNIVASDSVNGKMT-----LSUKDVPWDQALDVMQAR 402  
Db 99 TINMKDAEIGDFIEQVSSISGQTFYVDPVRKGVTVWSQARLSAEV--YOLFSLVLATH 156  
QY 403 NLDMSQOGNIVNIAPRDELLAKDKAFLOAEKDIA----- 437  
Db 157 GYAVLPQGDHARIVENME--ARQDA---AQTVRDGPGSLETRVVQAHQTSVAELIPMR 211  
QY 438 -----LGALYSQNFQKYNVEFRSILR----- 461  
Db 212 PLVPAHGHAAVPSANALIVSDRRSNIERIEAIVRLDRAGEHDYSIYDMRHAWVAEIAE 271  
QY 462 -LDN-----ADTTGNRNTLV----- 475  
Db 272 VLDRSVTPAAGKSAATVQVLADSRNSRLVLGGPQARLRLRLAQSLDVPSSRSANSRVI 331  
QY 476 -----SGRGS-----VLIDPATNTLIVTDRSVIEKF 502  
Db 332 RMRHGDAKTLAATLGEIGESLHGERGQDGRSGKRGKLLVRADESALVALILADPEDVGLL 391  
QY 503 RKLIDELVPAQOVMEIARIVEAADFSRDLGVKFGATGKKLKNDSIAFGWG-VNSGFG 561  
Db 392 EDIVQQLDVPRAQLLVEAAIVELSGEIGDALGVOMA-----LRSGHVAGGAGFADSGLS 445  
QY 562 GDDKWAETKINLPITAAANSISLVRAISSGALNLELSASLSKTKTLANPRVLTQNRK 621  
Db 446 IGTLLGALQAGKPPAELPDGALVGLSRDFGALVTALSRN--SRSNLSLTFSLTLTDNQ 502  
QY 622 EAKIESGVEIPEF---TVTSIANGSSNTTELKAVGLGT--VTPNITPDGQIIMTV--KI 674  
Db 503 KAEILVGNVFPFGTSYTTASGSSNPFTTVERKDIQVTLKVTPHIGEDRMLRIEIEEI 562

QY 675 NKDSPAQCAGNQTILCTSKNLNTQAMVNGGTLIVGGIYEEDNGNTLTQVPLLGIDIPV 734  
 Db 563 SSAPTATLAAKAVDLTKRSIKSTVLADGGQVIVLGGIQQDLQSRVPLLGIDIPG 622  
 QY 735 IGLFKTRGKTKDRRELLIFITPRIMGTA 763  
 Db 623 VGRFLRSSRTRVKRNLMLVFLRPSIVRDA 651

RESULT 9

US-09-252-991A-22297  
 ; Sequence 22297, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252.991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 22297  
 ; LENGTH: 995  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-22297

Query Match 6.9%; Score 264; DB 4; Length 995;  
 Best Local Similarity 26.5%; Pred. No. 1.4e-12;  
 Matches 89; Conservative 70; Mismatches 117; Indels 60; Gaps 15;  
 QY 469 GNRNLTLSVGRG-SVLIDPATNLIVTDRSVEIKRKLDELDPVPAQVNMIEARIVEAAD 527  
 Db 627 GEENSAFAGGVTVQADATNTLLISAPELYRNLREVIDLLDQRAQVIESLIVEVSE 686  
 QY 528 GFSDGLGVKGATGKKLNDTSFGVNSGFGDDKWAETKINLPITAAANSIS-IV 586  
 Db 687 DDSSEFQIQWAGN-LLGNGVFG-GVNF-OSALN-TAGKNTIDVLP 729  
 QY 587 RAISSGALN-LELS-ASELSKTKLANPRVLTQNRKEAKIESGYEI 631  
 Db 730 KGLNIGLVGTVDIPGICKLIDLVARALKSRGTNVLTSTPNTLLDNEASIMVGTI 789  
 QY 632 PFTVTS-IANGSSTNT-ELKAVLGLTVTPNTPGQIMTVKI-----NKDS 678  
 Db 790 PFVSGQVYTDGGTSNPFQTIQREDVGLKUNIRPQISEGG-TVKLDVYQEVSSVDE 845  
 QY 679 PAQCASGNQITLCISTKNLTQAMVNGGTLIVGGIYEEDNGNTLTQVPLLGIDIPVIGNL 738  
 Db 846 RASTAAGVVT-NKRAIDTSILLDDQIMVLGGLLDQNVQDNTDGVGSLPGVGS 901  
 QY 739 FKTRGKTKDRRELLIFITPRIM-GTAGNSL-RY 769  
 Db 902 FRYQKRSTKTNLMVFLRPYIVRDAAGRSITLNR 937

RESULT 10

US-09-252-991A-29967  
 ; Sequence 29967, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252.991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 29967  
 ; LENGTH: 752  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-29967

Query Match 6.6%; Score 252.5; DB 4; Length 752;  
 Best Local Similarity 19.8%; Pred. No. 7.8e-12;  
 Matches 151; Conservative 106; Mismatches 262; Indels 243; Gaps 26;  
 QY 106 NKQGYNTEVRGNKVMIFINSDTTSAPAPVAKAAPAK-QQCRTYVQVRSIRIQ 164  
 Db 31 DRTGSYHLRIDQSV-LLLRQGDLL-LSPLEHAPLDPRQOQGLRLLRVSASWS 86  
 QY 165 TLYPGKTTAAAPFTESV-SSVAPFSPAKQQAASA 199  
 Db 87 RYYP-QAIVLDADGRLLOARLGLDGLPERLERALAPGGAR---GAGA 133  
 QY 200 KOQTAAPAKQQTAA-PAKQQAAPAKQTNIDFRKDGKNAGIIE 241  
 Db 134 GAGTAAPRVAAGSGVASLMRRLLIGLLALLPGAVLRAQPLDWPSPDYVAQGESLRD 193  
 QY 242 LAALGFAGQPDIS-QQHDHIIIVTLKNTLPTTLQSLDVA- - - - -PKT 284  
 Db 194 VLA-NFGANYDASVIVSKVNDQVSGRFDLESPOAFLQMASLYNLGHWYDGTLYVFKT 252  
 QY 285 PVQKVTLKRLNN-DTQLIITTAGNWE-LVNKSAAPGYF-TF 322  
 Db 253 TEMQSLVRLVLEQVGEAEKRALTAAGIWEPRGWRADPSGRLVHS-GPGRYLELVEQTA 311  
 QY 323 QVLPKKQNLSEG-GVNNAPKFTT-GRKISLDQDVEIRTLQILAKE-SGMNIV 373  
 Db 312 QYLEQCYTLRSEKTDGLSVEIFPLRYAVAEDEKRIEYDEIAPGIAISLRVJSDANVV 371  
 QY 374 ASDSVNGKMTLSLKDVPWDQALDLVMOARNLDMROGNIIVNIAPRDELLAKDKAFLOAEK 433  
 Db 372 AVGDPEPKLR-PGQSSHAVVQA- - - - - 393  
 QY 434 DIADLALYQNFQKLYKNVEEFSIRILDNADTTGNRTLVSGRGSVLIDPATNLTIVT 493  
 Db 394 - - - - -EPLSNVAVVR 403  
 QY 494 DTRSVIEKFKLIDELVPAQVMI-EARIVE-AADGFSRDLGVKFGATGKKLKL- 546  
 Db 404 DHKDRLPYRRLIEALDRPSARIEVGLSIIDINAENLAQLGVDMWSAGIRLGNKSIQIRT 463  
 QY 547 -NDTSFAGVNSGFGDDKWAETKINLPITAAANSISLVRAISSGALNLELSASELS 605  
 Db 464 TQDSEEGGAGNAGV-SLVDSRGLDFLLAKVTLLSQG 502  
 QY 606 KTKTLANPRVLTQNRKEAKIESG-YEIPFTVTSIANGSSSTNTLKKAVLQ-LTVTPN 661  
 Db 503 QAQIGSRPTLLTQENTQAVLDQSEYTYVVTGERVA- - - - -ELKAITGTMLKXTPR 554  
 QY 662 ITPDG-QIIMTVKINKDSPACASGNQITLCISTKNLTQAMVNGGTLIVGGIYEED 718  
 Db 555 VVTLDGTPEISLSLHIEDGSQKPNAGLDKIPNTINRTVIDTITARVGHQSLLIGGIYRDE 614  
 QY 719 NGNTLTQVPLGDIPIVIGNLFKTRGKTKDRRELLIFITPRIM 760  
 Db 615 LSQSORKVPWLGDIPYLGALFRTTADTVRRSVRLFLIEPRLI 656

RESULT 11

US-09-198-452A-744  
 ; Sequence 744, Application US/09198452A  
 ; Patent No. 6559294  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Griffais, R.  
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragme

; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention of infection, and treatment of infection  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 744  
; LENGTH: 561  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-744

Query Match 6.1%; Score 233; DB 4; Length 561;  
Best Local Similarity 22.3%; Pred. No. 1.9e-10;  
Matches 106; Conservative 91; Mismatches 103; Indels 96; Gaps 17;  
QY 349 SLDQDVEIR-----TILQILAKES-----GMN---IVASDSVNGKMTLS 385  
DB 66 SVDTEYEVKYPANPAALVSYQDVLGTLAEDDAFQMFQGTNKFVYSSPRLANKAEQL 125  
QY 386 LKQVPMQALDLVMOARNLDMRQGNIVNIAPRDELLAKDKAFQAEKDIADLGALYSQN 445  
DB 126 LK-----SLDVPMAHTLD-----DPASTALALGGTTSFK-----SLRFFM 163  
QY 446 FQKYKXVEFRSILR---LDNADTGNRLTVSGRSLVIDPATNLTIVDTRSVIEKF 502  
DB 164 YKLAYGVEVIANALQDIGNLYVTTAMDEDFINTLSIQWLEVNNSIVIGNQGNVDRV 223  
QY 503 RKLDELDPVPAQVMIEARIVEADGSRDLGVKFGATGKKLKNDSAFSGWV---NSGF 560  
DB 224 IGLLNGDLPPKQYVIEVLIDTSLKSWDPGVQWVALGDEQSK---VAYASGLLNNIGI 280  
QY 561 GGDKWGAETKINLP-----ITAAANSISLVRAISSG----- 592  
DB 281 -----ATPTKATVPPTGPGSIPPTPGQLTGFSMLNSSSAFGLGIIGNVLSHKGS 334  
QY 593 --ALNLELSASESKYKTLANPRVLQNRKEAKIESGYRPP--TVTSIANGSSST-NT 647  
DB 335 FLTGLGLSALDQDGTIVVUNPRIMAQDTQCAEFFVQGVYPTQTTNTIIQETGTVQNI 394  
QY 648 ELKXAVLGLTVPNTDQGIIMTKINKD-SPAQCASGNQITLCISTKNLNTQAMVYNG 706  
DB 395 DYEDIGNLVVSTVAPNN--VWILQIEQITSELHSASGLTPTVTKT-YAATRLQIPDG 451  
QY 707 GTLVGGIYEEDNGLTKVLLGDIPVIGNLFTKRGKKTDRELLIFITPRINGT 762  
DB 452 CFLVMSGHIRDKTKVWSGVPLNLSIPLIRGLFSTIDQQRKNIMFIKPKVISS 507

RESULT 12  
US-09-198-452A-874  
; Sequence 874, Application US/09198452A  
; Patent No. 6559294  
; GENERAL INFORMATION:  
; APPLICANT: Griffiths, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention of infection, and treatment of infection  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 874  
; LENGTH: 754  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-874

Query Match 5.5%; Score 212.5; DB 4; Length 754;  
Best Local Similarity 21.3%; Pred. No. 1.4e-08;  
Matches 111; Conservative 82; Mismatches 217; Indels 111; Gaps 16;  
QY 287 QKVLKRLNNDTQLIITAGNWLWNKSAARGYTFQVLPPKQNLSESGVNNAPKTTGR 346

DB 234 QHVLKFFINPETHVDVIVAGRWVIFSGAGEVG-----ELLKIYNFVQSESIQREYRVIPLT 289  
QY 347 KISLDFQDVEIRITQILAKESGMNIVASDSVNGKMTLSLKDVPWDQ-----ALDL 397  
DB 290 KI-----DPGEMISILNAAAFREDLTAKDVSSES-----LGLRVVPLQYQGRSLFSGTAAL 339  
QY 398 VMOARNLDMRQGNIVNIAPRDELLAKDKAFQAEKDIADLGALYSQNQLKXKXVEFR 457  
DB 340 VOQALTILRELEEGEN--PTDKTVF---WYNNKXSDPQELAALLSQ-----VHDVFSGE 389  
QY 458 SILRLDNADTTGNR-----NTLVSGR-----GSLVIDPATNLTIVDTRSVIEKF 502  
DB 390 NKASVGAADGCGSQLNASIQIDTTVSSSAKDGSKVYGNFIADSKTGTLLIMVYKEVLPRI 449  
QY 503 RKLDELDPVPAQVMIEARIVEADGSRDLGVKFGATGKKLKNDSAFSGWVNGSGFG 562  
DB 450 QMLLKKLDLPKQVRIEVLFFERKLAHQKSGNLRLRGEVCKKGC----- 497  
QY 563 DDKWAETKINLPITAAANSISLVRAISSGALNLSASELS-----KTKTLANPR 614  
DB 498 -----PSVSWAGGTGILEFLFKGSTGSSIVPGYDLAYQFLMAQEDVRINASPS 545  
QY 615 VLTQNRKEAKIESGYEIPFTVTSIANGSSSTNTELKAVLGLTV--TPNIT---PDQIIL 669  
DB 546 VVTMNTQTPARIAVDEM-----SIAVSDKKAQVNRAGYGMIXMLPVINVEEDGKSY 600  
QY 670 MTV-----KINKDSPAQCASGNQITLCISTKNLNTQAMVYNGGTLIVGGIYEEDN 719  
DB 601 ILELTDITDTTGNHDDRPD-----VTRNITNKVRIADGETVILGLRCKQM 649  
QY 720 GNTLTKVLLGDIPVIGNLFTKRGKKTDRELLIFITPRIM 760  
DB 650 SDSHDGIPFLGIPGIGKLFMGSSSTSDSLTEMFVITPKIL 690

RESULT 13  
US-09-252-991A-26461  
; Sequence 26461, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107195.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 26461  
; LENGTH: 594  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26461

Query Match 5.0%; Score 192.5; DB 4; Length 594;  
Best Local Similarity 19.3%; Pred. No. 3.9e-07;  
Matches 134; Conservative 119; Mismatches 269; Indels 173; Gaps 30;  
QY 135 ARPAAKAAAPAAKQKQCRVTVQVRSIRIQTLTPGKTTAAAPFTESVWSVSAFFSPAKQ 194  
DB 5 ALPGERQPAAPVDHOGAQLMR-----APLKNLLPCLLIP-----AL 42  
QY 195 AAASAKQQTAAAPAKQQTAAAPAKQQAAPAKQTNIDFRKDGKAGIIEALGAGOPDIS 254  
DB 43 ASCSVTRVNESADRVATADASTIAAQMNRTPD-RRD-----TVVFSDDKPVWS 91  
QY 255 QQHDHIVTLKXHTLPTTLQRLSDLVADFKTPQVKVTKRLNNDTQLIITAGNWLWNK 314  
DB 92 TKPLSVSHLSSDCIYV--WRPAGASLQEAQAEVI-----NQCHLAVSITP--DALNPA 142



QY 315 AARGYFTFQVLPKKQNLGSGVNAKPTFTGRKISLDFQDVEIRTIQILAKESGNVIVA 374  
Db 143 A-----FALPOQOR-----ASNAPPPIOGQ-----DNATWLPASVANGSLGA 182  
QY 375 SDSVNGKX-----TSLKDVDPDQA-----LDLVMOARNLDMRQ-----QGNIVN 414  
Db 183 GSGISGFSVGSYGRSLYNINIKWNGKVSGLDLIAARAGVSRVNPTEKRVFFYLDLTRFR 242  
QY 415 IAPDELLAKDKAFLOAEKDIA-----DLGALYSNFOFKYKNVE 454  
Db 243 IYAFDDVNTVDSTVRSGMTTAAGISGCGSGSTGONGSSGSGSGSKQTSSSELK----- 297  
QY 455 EFRSILRLDNADTTGNRNTLVS--GRGSVLIDPATNTLIVTDRSVIEKFKLID-ELDVP 512  
Db 298 --TSIL-----SDIENSINMLTPSGRMSLSRATGTLTVDTRBEVLNRVQOVLNRENESI 351  
QY 513 AQQVMIEARIVEAADGSRDLGVAFGATGKKLKNDSAFSGWGVNSFGGDDKWKGAETKI 572  
Db 352 TKQVLLNVNLSVALTKDQJGIDNVLV--YKSLNN-----KWGIGLKN 393  
QY 573 NLP---ITAAANSISLVRATISS--GALNLELSASELSKTKTLANPVLTONRKEAKIES 627  
Db 394 TMPGIDQSAISGSVILDTANSAGSKAMVQALAAQQGRVSTVRSPSVTTLNLSAPIQI 453  
QY 628 G-----YEIPFTVTSIANGSGSTNTELKAV---LGLVTPNITPDGOIIMTVKINK----- 676  
Db 454 GRYSYASSQISNVAQVGSFTSL-IPGAVTSGYNMSSLPPVMSGEMLLKININMTSRP 512  
QY 677 -----DSPAQCASGNQITLCISTKNLNTQAMVNGGTLIVGGIYE--EDNGNTLTKV 726  
Db 513 TFEMOTSGSKAQFPS-----VDIQLFDQKVALRSGETLVLSGFDOTTEDTNKVGI-- 563  
QY 727 PLLGDPVIGNLFKTRGKKTDRRELLIFITPRIMG 761  
Db 564 ---GDAGFFG-LGGGLTRNTKREVIVLITPVVLG 594

RESULT 14  
US-09-252-991A-23140  
; Sequence 23140, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23140  
; LENGTH: 1246  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23140

Query Match  
Best Local Similarity 4.8%; Score 183.5; DB 4; Length 1246;  
Matches 90; Conservative 65; Mismatches 144; Indels 107; Gaps 17;  
QY 395 LQVMOARNLDM--RQGNVNIAPDELLAKDKAFLOAEKDIALGALYSNFOFKYK- 451  
Db 43 IELLAQAPVDVVGQGDRLAIVPIERLAIGD-----PKIADVQLDRGRFLVTGKE 95  
QY 452 -----NVEBFRSLRLDNADTTGNRNTLVSGRGSVLIDPATNTLIVTDRSVI 499  
Db 96 QGSTSLIWTGCGPEPLRSIVE-----VEGGSV-----DTRGAP 130  
QY 500 EKFRKLIDELDPVPAQVMIEARIVEAADGFSRDLGVKF-----GATGKKLKL 546

Db 131 AFTVGAEEEL---PNQVOTDIRFVEVSRSKLKQASTSFVRGGLNWLVPFSSGLGDIKNV 187  
QY 547 NDTSS---AFGWGVNSGFG---GDDKMGAEKTKINLPITAAANSISLVRATISSGALNLELS 599  
Db 188 ADGSGLGCTFGTG--SSGFNLIFGGGKM-----LSFMNAL----- 220  
QY 600 ABESLUSKTKTLANPVLTONRKEAKIESGIYIPFTVTSIANGSGSTNTELKAVLGLTWT 659  
Db 221 --EGSGFAYTTLARPSLVAMSGOSASFLAGGEPFPVPNGTK--DNVTYKEFGIRLTLT 276  
QY 660 PNITPDGOIIMTV--KINK-DSPAQCASGNQITLCISTKNLNTQAMVNGGTLIVGGIYE 716  
Db 277 PTVMNRRIALKAIEVSELDVSAIGQNGGVAVPALRVRRTDTSVMLADGGSFVIGLTS 336  
QY 717 ENGNTLTKVPLGLDIPVIGNLFKTRGKKTDR--RELLIFITPRIM 760  
Db 337 SNSVSNVDKFFLAG-ATRSRSVRSFRSTKVDDKDDRELLMIVTPHLV 381

RESULT 15  
US-09-206-942-49  
; Sequence 49, Application US/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; TITLE OF INVENTION: Molecular Weight Proteins  
; FILE REFERENCE: 1038-861 MIS:35  
; CURRENT APPLICATION NUMBER: US/09/206,942  
; CURRENT FILING DATE: 1998-12-08  
; EARLIER APPLICATION NUMBER: 09/167,568  
; EARLIER FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 49  
; LENGTH: 1073  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-206-942-49

Query Match  
Best Local Similarity 4.2%; Score 160.5; DB 4; Length 1073;  
Matches 125; Conservative 88; Mismatches 212; Indels 211; Gaps 25;  
QY 205 APAKQQTAAAPAKQAAAAAPAKQTNIDFRKDGKNAGIIEALALGAPAGQPDISQQHDIIVTL 264  
Db 8 APSAERT-----DTGEDVEYTGADT-----NHQKQNSKSTL 42  
QY 265 KNHTLPTTLORSLOVDADFKTPVQKVTLKRLNNDTQLIITTAGHWELVWKSAAFGYPTFQV 324  
Db 43 TMTLEGMLKREL-----FVNITARKIRVNSTINIGDSGLT 80  
QY 325 LPKKONLESQGVN-NAPKTTTGRKISLDFQD-VEIRTIQILAKESGMNIVASDSV--- 378  
Db 81 LYKKRKRSQGIQINKDITSTGSLTINSDDWDIHG--NITLGEGLNITSSDSVAFEG 138  
QY 379 -----NGKMTLS-----LKQVPWD---QALDLYMOARNLDMRQOGNI 412  
Db 139 GNGNKRSSASAQIIAAGTTILTGENTKTFRLNNVSLNGTNGLSIISTAGNLSHRLDGEI 198  
QY 413 VNIAPR---DELLAKDKAFLOAEKDIALGALYSNFOFKYKVEEPRSLRLDNADT-- 467  
Db 199 -NVSNVTINQTTQONTQNIYHWKASD-----SYMNVTSFN--LREDSKFTFI 241  
QY 468 ---TCNRTNLVSGRGSVLIDPATNTLIVTDRSVIEKFKLIDELDPVPAQVMIEARIVE 524  
Db 242 KYVNSARNGDVRGRSF-----AGVIF 262  
QY 525 AADGFSRDLGVKFATGCKKLNKNTSAFGWGVNSFGGDDKWKGAETKINLPITAAAN-SI 583

Db 263 NAKGLTTSFNVKGVSTVDFKLK-----PNSGYNQKRIPIQFQSNISVSGGGRVNI 313  
Qy 584 SLVRAISSGALNLELSASELSKTKTL-----ANPVLVLTQ- 618  
Db 314 NTLANLTGGGVEIR-SSSINVSGSTLSMTAQARDNAFEITKDLVINASNSLSIIQON 372  
Qy 619 -----NRKEAKIESGYEIPFTVTSIANGSGSTNTELKAVLGLTVTPNITPDGQIIMTVK 673  
Db 373 DGFNNOKANAINSKYNTIQQGNVTLGGQNSSSTITGSV-----NIGANAVTLOAH 425  
Qy 674 INKDSPAQACASGNQTI-----LCISTKNLNTQAMVENG- 720  
Db 426 NGNDRNKKLTFGNVSEGEGLRVGASANINNNLSVKSGAKFAETNDNLNITGTF-TNNG 484  
Qy 721 NTLTKV-----PLLGDIPIVGNLFKTRGKKTDRRELL 752  
Db 485 TSIIIDVKKGAACKGNITNDGNLNIITNAKNGQKSVI 520

## RESULT 16

US-09-206-942-47  
; Sequence 47, Application US/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; TITLE OF INVENTION: Molecular Weight Proteins  
; FILE REFERENCE: 1038-861 MS:jb  
; CURRENT APPLICATION NUMBER: US/09/206,942  
; CURRENT FILING DATE: 1998-12-08  
; EARLIER APPLICATION NUMBER: 09/167,569  
; EARLIER FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 47  
; LENGTH: 1079  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-206-942-47

Query Match 4.2%; Score 160.5; DB 4; Length 1079;  
Best Local Similarity 19.7%; Pred. No. 0.00039;  
Matches 125; Conservative 88; Mismatches 212; Indels 211; Gaps 25;  
Qy 205 APAKQOAPAKQOAAAPAKQTNIDFRKQGNAGIIEALALGFAGQPDISQQHDIHIVTL 264  
Db 14 APSAERT-----DTGEDVEYTGTDI-----NHQKQNSKSTL 48  
Qy 265 KNHTLPTLQSLDADFKTPQVKVTLKRLNNDTQLIITTAGHWELVNKSAAPGYFTFQV 324  
Db 49 TNTLEGNLRGL-----FNITARKIRVNSTINIGDSGHLT 86  
Qy 325 LPKKQNLSEGGVN-NAPKTTGTRKISLDFO-VEIRTILOILAKESGMNIVASDSV- 378  
Db 87 LYKKRNRSQGIQINKDITSTGGSLTINSDDVDIHG--NITLGEGLNITSSDSVAFEG 144  
Qy 379 -----NGKMTLS-----LKQVWD---QALDLYVQARNLDMRQGNH 412  
Db 145 GNGNKGSSASAQIIAGTTLTGKNTFRNLNVSLNGTONGLSIISTASNSHRLDGEI 204  
Qy 413 VNTAPR---DELLAKDKAFLOAEKDIADLGALYSQNFQLYKXVBEFRSILRLDNADT-- 467  
Db 205 -NVSGNVTINQTOQNIEYWKASSD-----SYMNVTSEF--LREDSKFTFI 247  
Qy 468 ---TGNRNTLVSGRGSVLIDPAINTLIVTDRSVIEKFKLIDELDVPAQNMIEARIYE 524  
Db 248 KYNSARNGDVRGRSF-----AGVIF 268  
Qy 525 AADGFSRDLGVKFGATGKKLNDTSAFGWGNSGFGDDKWAETKINLPITAAAN-SI 583  
Db 269 NAKGLTTSFNVKGVSTVDFKLK-----PNSGYNQKRIPIQFQSNISVSGGGRVNI 319

## RESULT 17

US-09-328-352-5723  
; Sequence 5723, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 5723  
; LENGTH: 1088  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-5723

Query Match 4.1%; Score 157.5; DB 4; Length 1088;

Best Local Similarity 21.3%; Pred. No. 0.00069;  
Matches 191; Conservative 111; Mismatches 336; Indels 259; Gaps 45;

Qy 14 VATAFQTASAGNITDIKVSILPNK-----QKIVKVSFDEIVNPTGF-VTSSPARIALD 67  
Db 249 IQCAINRNQAAGLPDIDYWSLKLQETMNVVPRFLAVA--QIKNPRAYGSLPIPIANRPH 306  
Qy 68 FEQTGISVQDQVLEVADPL-LSKISAAONSRRRL-VLNLNKPGQ----- 110  
Db 307 PREVTLSA-----PLSLNEIASVTGLSRAELYALNPGYRGETVDPSPMRILIPAD 357  
Qy 111 ----YNTVEVRGNKV-----WIFI-----NESDVTVSAPARPAKAAAPAKQ 150  
Db 358 IPSVDNKLKGMKAGSSGWNASVTSKPTTTTSTSVTVTTSTPAQVPRSPFPAKTS 417  
Qy 151 GCRVTYVRSIRIOTLYPGKTTAAAPFTESVVSAPSPAKQQAASAK----- 200  
Db 418 S-----SSVTVKTATPRGSDALAAFAASADVPSAPRIPVAVTPAANNIKPVRTPEPISA 470  
Qy 201 -----QOATAAPAKQOATAAPA-KQAAAAPAKQ-----TNIDFRKQGNAGI 239  
Db 471 TEREXILAAVRAEBKETVDQALEPQATQAKDQVVAELKALAPOGTEIVDPYDK- 526  
Qy 240 IELALGFAGQPDISQQHDIHIVTLKHNLTPTTL-----QRSLDVADFPTVQ 287  
Db 527 IKLTAIQ-TSQSVABQOQKEV---SKGPAYPKTLAEDATLANSEDAQRNKKPKYIKTDTD 582  
Qy 288 KYTL-----KRLN-----NDTQLIITTAG--KNELVNKSAAPGYFTFQVLPKKQNLSEGGV 336  
Db 583 VVVVQPKGRSTYTVQPGDTLAVTAMKGVNWRDVAK-----WNQIDPEK----- 627  
Qy 337 NNAPKTFTRKISL-DFQDVEIRTILOILAKESGMNIVASDSVNG---KMTLSLKDVPWD 392  
Db 628 ---TLFVGTSLYLYDAKPOAEITAKSAAPDVVVVQANDSLTGAVNQFNLSVKQLAEY 683  
Qy 393 QAL----DLVMQARNLDMRQ-QGN-IVNTAPR-----DELLAKDKAFLOAE--KDIADLG 439



```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-038-682-2

Query Match      4.0%; Score 155.5; DB 1; Length 1536;
Best Local Similarity 19.1%; Pred. No. 0.0017;
Matches 146; Conservative 123; Mismatches 284; Indels 213; Gaps 41;

QY 25 GNITDIKVSLLPNKQKIVKVSFDEKIVNPTGVTSSPARIALDFEQTGISMDQVLEYAD 84
Db 567 GNI-----NITAKQDIAFEKGSNQVITGGTITSG-----NQKGRFNN-----605

QY 85 PLLSKISAAQNSRRARLVNLNKPQ---QYNTVEVGNKVMWIFINSDDOTVSAPAPAKVA 141
Db 606 -----VSLNGTSGLOFTTK-RTNK-YAITNKFEGLTNSGK--VNI 643

QY 142 APAAPAKQOQC-----RTVYQVRSIRIOTLYPGKTTAAAPFTESVVSAPFSPAKQQA 196
Db 644 SMVLPKNSGYDKFKGRTYNNLTSNVSESGEFNLIDSRGSDSAGTLTOPYN-----696

QY 197 ASAKQOQTAAPAKQOQTAAPAKQOTNIDFRKDGKNAGIIELAAALGFA-----248
Db 697 -----LNGISFNKDTTFNVERNA-----RVNFDIKAP-----TFGSRKAVTEIEGNVT 743

QY 249 ---GQPDISOQHDHIIIVTLKNHTLPTLQRS---LDVAD-----FKTPVQKVTLKRLNNDT 298
Db 744 SGGGSVDFT-----LLASSNVQTPGVVINSKYFNVTSGSSLPFKTSKTKTGFSEIKDL 798

QY 299 QLIITTAGNWLNVKSAAPGYFTFQVLPKQK---NLESGGVNNAKPTFTGRKISLDFQ-DYE 356
Db 799 TL-NATGGNITLLQVEGTDMIGKGVAKKNITFEQGN-----TFGSRKAVTEIEGNVT 852

QY 357 IRTILOILAKESGMNIVASDSVNGKMTLSLKDVPDQALDLMVQAENLDMRQGNIVNIA 416
Db 853 IN-----NNANVTLLIGSDFDNHOKPLTIKK-----DVTNSGNL---TAGGNIVNIA 896

QY 417 PRDELLAKDAFLQAEKDIA-DLGALYSQNFQKLYKQVVEFRSILRLDNADTTGNRNTIV 475
Db 897 --GNLTVESNANFKAITNFTFNVGGLF-----DNKGSNISI 931

Query Match      4.0%; Score 155.5; DB 1; Length 1536;
Best Local Similarity 19.1%; Pred. No. 0.0017;
Matches 146; Conservative 123; Mismatches 284; Indels 213; Gaps 41;

QY 25 GNITDIKVSLLPNKQKIVKVSFDEKIVNPTGVTSSPARIALDFEQTGISMDQVLEYAD 84
Db 567 GNI-----NITAKQDIAFEKGSNQVITGGTITSG-----NQKGRFNN-----605

QY 85 PLLSKISAAQNSRRARLVNLNKPQ---QYNTVEVGNKVMWIFINSDDOTVSAPAPAKVA 141
Db 606 -----VSLNGTSGLOFTTK-RTNK-YAITNKFEGLTNSGK--VNI 643

QY 142 APAAPAKQOQC-----RTVYQVRSIRIOTLYPGKTTAAAPFTESVVSAPFSPAKQQA 196
Db 644 SMVLPKNSGYDKFKGRTYNNLTSNVSESGEFNLIDSRGSDSAGTLTOPYN-----696

QY 197 ASAKQOQTAAPAKQOQTAAPAKQOTNIDFRKDGKNAGIIELAAALGFA-----248
Db 697 -----LNGISFNKDTTFNVERNA-----RVNFDIKAP-----TFGSRKAVTEIEGNVT 743

QY 249 ---GQPDISOQHDHIIIVTLKNHTLPTLQRS---LDVAD-----FKTPVQKVTLKRLNNDT 298
Db 744 SGGGSVDFT-----LLASSNVQTPGVVINSKYFNVTSGSSLPFKTSKTKTGFSEIKDL 798

QY 299 QLIITTAGNWLNVKSAAPGYFTFQVLPKQK---NLESGGVNNAKPTFTGRKISLDFQ-DYE 356
Db 799 TL-NATGGNITLLQVEGTDMIGKGVAKKNITFEQGN-----TFGSRKAVTEIEGNVT 852

QY 357 IRTILOILAKESGMNIVASDSVNGKMTLSLKDVPDQALDLMVQAENLDMRQGNIVNIA 416
Db 853 IN-----NNANVTLLIGSDFDNHOKPLTIKK-----DVTNSGNL---TAGGNIVNIA 896

QY 417 PRDELLAKDAFLQAEKDIA-DLGALYSQNFQKLYKQVVEFRSILRLDNADTTGNRNTIV 475
Db 897 --GNLTVESNANFKAITNFTFNVGGLF-----DNKGSNISI 931

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-302-832-2

Query Match      4.0%; Score 155.5; DB 1; Length 1536;
Best Local Similarity 19.1%; Pred. No. 0.0017;
Matches 146; Conservative 123; Mismatches 284; Indels 213; Gaps 41;

QY 25 GNITDIKVSLLPNKQKIVKVSFDEKIVNPTGVTSSPARIALDFEQTGISMDQVLEYAD 84
Db 567 GNI-----NITAKQDIAFEKGSNQVITGGTITSG-----NQKGRFNN-----605

QY 85 PLLSKISAAQNSRRARLVNLNKPQ---QYNTVEVGNKVMWIFINSDDOTVSAPAPAKVA 141
Db 606 -----VSLNGTSGLOFTTK-RTNK-YAITNKFEGLTNSGK--VNI 643

QY 142 APAAPAKQOQC-----RTVYQVRSIRIOTLYPGKTTAAAPFTESVVSAPFSPAKQQA 196
Db 644 SMVLPKNSGYDKFKGRTYNNLTSNVSESGEFNLIDSRGSDSAGTLTOPYN-----696

QY 197 ASAKQOQTAAPAKQOQTAAPAKQOTNIDFRKDGKNAGIIELAAALGFA-----248
Db 697 -----LNGISFNKDTTFNVERNA-----RVNFDIKAP-----TFGSRKAVTEIEGNVT 743

QY 249 ---GQPDISOQHDHIIIVTLKNHTLPTLQRS---LDVAD-----FKTPVQKVTLKRLNNDT 298
Db 744 SGGGSVDFT-----LLASSNVQTPGVVINSKYFNVTSGSSLPFKTSKTKTGFSEIKDL 798

QY 299 QLIITTAGNWLNVKSAAPGYFTFQVLPKQK---NLESGGVNNAKPTFTGRKISLDFQ-DYE 356
Db 799 TL-NATGGNITLLQVEGTDMIGKGVAKKNITFEQGN-----TFGSRKAVTEIEGNVT 852

QY 357 IRTILOILAKESGMNIVASDSVNGKMTLSLKDVPDQALDLMVQAENLDMRQGNIVNIA 416
Db 853 IN-----NNANVTLLIGSDFDNHOKPLTIKK-----DVTNSGNL---TAGGNIVNIA 896

QY 417 PRDELLAKDAFLQAEKDIA-DLGALYSQNFQKLYKQVVEFRSILRLDNADTTGNRNTIV 475
Db 897 --GNLTVESNANFKAITNFTFNVGGLF-----DNKGSNISI 931

; Sequence 2, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
```

QY 476 SGRGVL--IDPATNTLIVTDRSVIEKPKLID-----ELDVPAQ 514  
DQ 932 AKGAREFKDINSKSLGTTNSST---YRTIISGNITKNGDLNITNEGSDEMIGGD 988  
QY 515 QVMIEARIVEAAD--GFSRDLGVKFGATGKKLKNLNTSFAFGWNSGFGGDDKWAETKI 572  
DQ 989 VSQKEGNLTISDDKINIKITKAGVDGENSDSDATN-----1026  
QY 573 NLPITAAANSISLVRAIS--SGALNLELSASE---SLSKTKTL--ANPRVLTNR--KEAK 624  
DQ 1027 NANLTITKELKLTQDLNLSGFNAEITAKDGSDLTIGNTSADGTTNAKKVTNFQVDSK 1086  
QY 625 IES--GYEIPFTVTSIANGGSSTNTE-LKAVLGLTV--TPNITPDGQIIMTVKINKOSPAQ 681  
DQ 1087 ISADGHKVTILH--SKVETSGSNNTEDSSDNNAGLTIDAKNVTVNNITS-----HKAVSIS 1141  
QY 682 CASGNQTLICISTKNL---NTQAVENGGLTIVGGIYEEDNGNTLT 724  
DQ 1142 ATSGETITKTGTATTGNVEITAQTGS--ILGGIESSSGSVTLT 1185

RESULT 21  
US-08-530-198-2  
; Sequence 2, Application US/08530198  
; Patent No. 5869065  
; GENERAL INFORMATION:  
; APPLICANT: BARENKAMP, STEPHEN J  
; APPLICANT: ST. GEME III, JOSEPH W  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/530,198  
; FILING DATE: 13-DEC-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BERKSTRESSER, JERRY W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: JWB-1186  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1536 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-530-198-2

Query Match 4.0%; Score 155.5; DB 2; Length 1536;  
Best Local Similarity 19.1%; Pred. No. 0.0017;  
Matches 146; Conservative 123; Mismatches 284; Indels 213; Gaps 41;

QY 25 GNITDKVSLPNKQKLVKVSFDEKIVNPTGFTVSSPARIALDFEQTGISMDQVLEYAD 84  
DQ 567 GNI-----NITAKQDIAFKSGNQVITGGQTTISG-----NOKGFRFNN-----605  
QY 85 PLLSKISAAQNSRRARLVNLNPKFG---QYNTVEVRGNKWNIFINESDDTVSAPARPAVKA 141

DB 606 -----VSLNGTSGSLQFTTK--RTNK--YAITNKFEGTLNISK--VNI 643  
QY 142 APAAPAKQGC-----RTVQVRSIRIOTLYPGKTTAAAPFTESVVSAPSPAKQAAA 196  
DQ 644 SWLFPNMSGYDXFKGRYVNLTLNLSVSESGEFLNLTIDSRGSDSACTLTQPVN-----696  
QY 197 ASAKOCTAAPAKQQAAPAKQQAAPAKQTNIDFKDGNAGIIEAALGEA-----248  
DQ 697 -----LNGISFNKDTTFNVERNA-----RVNFDIKAP-----IGINKYSSLNTASFNIGNISV 743  
QY 249 ---GQPDLSQOHDHIIIVLQHTLPTTLORS--LDVAD-----FKTPVQKVTLKLNNDT 298  
DQ 744 SGGGSVDFT-----LLASSNVQTFGVVINSKYFNVSTGSSLRFTSGSTKTGFSIEKDL 798  
QY 299 QLIITTAGNWEIWNKSAAPGYFTFQVLPKKQ--NLESGGVNNAKPTFTGRKISLDQ--DYE 356  
DQ 799 TL--NATGGNITLLQVEGTGMIGKIVAKNITFECCNI-----TFGGRKAVTEIEGNT 852  
QY 357 IRTILQILAKESGMNIVASDSVNGKNTLSLKDVPDQALDLVQARNLDMRQGNIVNIA 416  
DQ 853 IN-----NNANVTLIGSDFDNHQKPLTIKK-----DVIINSGNL--TAGGNIVNIA 896  
QY 417 PRDELLAKDAFLQAEKQIA--DLGALYSONFOLKYKNVEEPRSIILDLNADTTGNRNTLV 475  
DQ 897 --GNLTVESNANFKAITNFTFNVGGIF-----DNKGNISNI 931  
QY 476 SGRGVL--IDPATNTLIVTDRSVIEKPKLID-----ELDVPAQ 514  
DQ 932 AKGAREFKDINSKSLGTTNSST---YRTIISGNITKNGDLNITNEGSDEMIGGD 988  
QY 515 QVMIEARIVEAAD--GFSRDLGVKFGATGKKLKNLNTSFAFGWNSGFGGDDKWAETKI 572  
DQ 989 VSQKEGNLTISDDKINIKITKAGVDGENSDSDATN-----1026  
QY 573 NLPITAAANSISLVRAIS--SGALNLELSASE---SLSKTKTL--ANPRVLTNR--KEAK 624  
DQ 1027 NANLTITKELKLTQDLNLSGFNAEITAKDGSDLTIGNTSADGTTNAKKVTNFQVDSK 1086  
QY 625 IES--GYEIPFTVTSIANGGSSTNTE-LKAVLGLTV--TPNITPDGQIIMTVKINKOSPAQ 681  
DQ 1087 ISADGHKVTILH--SKVETSGSNNTEDSSDNNAGLTIDAKNVTVNNITS-----HKAVSIS 1141  
QY 682 CASGNQTLICISTKNL---NTQAVENGGLTIVGGIYEEDNGNTLT 724  
DQ 1142 ATSGETITKTGTATTGNVEITAQTGS--ILGGIESSSGSVTLT 1185

RESULT 22  
US-08-469-880-2  
; Sequence 2, Application US/08469880  
; Patent No. 5876733  
; GENERAL INFORMATION:  
; APPLICANT: BARENKAMP, STEPHEN J.  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/469,880  
; FILING DATE: 06-JUN-1995

CLASSIFICATION: 435  
PRIOR APPLICATION DATA: GB 9205704.1  
FILING DATE: 16-MAR-1992  
PRIOR APPLICATION DATA: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
APPLICATION DATA: US 08/302,832  
FILING DATE: 16-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-516 MIS:VG  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-469-880-2

Query Match  
Best Local Similarity 19.1%; Pred. No. 0.0017; Indels 213; Gaps 41;  
Matches 146; Conservative 123; Mismatches 284; Indels 213; Gaps 41;

QY 25 GNITDIKVSLLPNKQKIVKVSFDEIVNPTGFTVSSPARIALDFEQTGISMDDQVLEYAD 84  
DB 567 GNI-----NITAKQDIAFEKGSNQVITGGTITSG-----NQKGFRENN-----605

QY 85 PLLSKISAQNSRRARLVNLNKPQ---QYNTVRGNKQVWIFINESDDTVSAPARPAVKA 141  
DB 606 -----VSLNGTSGLOFTTK-RTNK-YAITNKEGTLNLSGK--VNI 643

QY 142 APAAPAKQGC-----RTVQVRSIRIQTLYPGKTTAAAPFTSVSVGAPFSPAKQQA 196  
DB 644 SMVLPRKESYDFKGTYNWLSLNVSESGENLIDSRGSDSAGTLTQPN-----696

QY 197 ASAKQQTAAPAKQTAAPAKQAAAQAKQNIIDFRKDGKNAGHIELAALGFA-----248  
DB 697 -----LNGISFNKDTTFNVERNA-----RVNFDIKAP---IGINKYSSLNYSFNGNISV 743

QY 249 ---GQPDISOQHDHIIIVTLKNHPTTLTQRS--LDVAD-----FKTPVQKVTLLKLNNT 298  
DB 744 SGGSDVDT-----LLASSNVQTPGVVINSKYFNSTGSLRFKTSGSTKTGFSIEKDL 798

QY 299 QLIITTAGNWLNVKSAAPGYFTFQVLPKQ-NLESGVNNAPKTTGGRKISLDFO-DVE 356  
DB 799 TL-NATGNTLLQVEGTGDMIGKIVAKKNITFEQNI-----TFGRKAVTEIEGNTV 852

QY 357 IRTILQILAKESGMNIVASDVNGKWTLSLKDVPWQALDLVMAQRLDMRQOQNIIVNTA 416  
DB 853 IN-----NNANVTIGSFDNHQKPLTIKK-----DVIINSGL--TAGGNIVNTA 896

QY 417 PRDELLAKRAFLQAEKDIA-DLGALYSQNFQKLYKNVEEFRSILRLDNADITGNENTLV 475  
DB 897 --GNLTVESNANFKAITNFTFNVGGLF-----DNKGNISNI 931

QY 476 SGRGSVL--TDPATNLIVDTSRVIEKFKRLID-----ELDVPQAQ 514  
DB 932 AKGARFKQIDNKNLSITNSST---YRTIISGNITNKNGLNTNMGSDTFEMQIGD 988

QY 515 QVMIEARIVEAAD--GFSRDLGVKFGATGKKLNDTISAFGWVNSGFGDDKRWGAETKI 572  
DB 989 VSQKEGNLTISDKINITKQITIKAGVDGENSDSDATN-----1026

QY 573 NLPITAAANSISLVRALS-SCANLELSASE-----SLSKTKTL--ANPRVLTQNR-KEAK 624  
DB 1027 NANLTIKTSLKLTQDLNLSGFKNAEITAKDGSDLTIGNTNSADGTNNAKVKVTFNQVKDSK 1086

QY 625 IES-GYEIPFTVTSIANGSGSSTNTE-LKKAVLGLTV-TPNITPDQOIIMTVKINKDSPAQ 681  
DB 1087 ISADGHKVTLLH-SKVETSGSNNNTEDSSDNNAGLTIDAKNVTVNNITS-----HKAVSIS 1141

QY 682 CASGNQITLCISTKNL---NTQAMVENGCTLIVGCIYEEDNGNTLT 724  
DB 1142 ATSGEITTKTGTATTNATTGNVEITAQTGS--ILGGIESSSGSVTLT 1185

RESULT 23  
US-08-728-470-2  
Sequence 2, Application US/08728470  
Patent No. 5928651  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,470  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-633  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE:  
US-08-728-470-2

Query Match  
Best Local Similarity 19.1%; Pred. No. 0.0017; Indels 213; Gaps 41;  
Matches 146; Conservative 123; Mismatches 284; Indels 213; Gaps 41;

QY 25 GNITDIKVSLLPNKQKIVKVSFDEIVNPTGFTVSSPARIALDFEQTGISMDDQVLEYAD 84  
DB 567 GNI-----NITAKQDIAFEKGSNQVITGGTITSG-----NQKGFRENN-----605

QY 85 PLLSKISAQNSRRARLVNLNKPQ---QYNTVRGNKQVWIFINESDDTVSAPARPAVKA 141  
DB 606 -----VSLNGTSGLOFTTK-RTNK-YAITNKEGTLNLSGK--VNI 643

```

PILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
PILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-617-697-2

Query Match 4.0%; Score 155.5; DB 2; Length 1536;
Best Local Similarity 19.1%; Pred. No. 0.0017;
Matches 146; Conservative 123; Mismatches 284; Indels 213; Gaps 41;

QY 25 GNITDIKUSSLPNKKIVKVSFDEKIVNPTGCVTSSPARIALDFEQTGISMQQQVLEYAD 84
Db 567 GNI-----NITAKODIAFEKSGSNVITGGQITSG-----NQKGRFPNN----- 605
QY 85 PLLSKISAAQNSRRARLVNLNKKPG---QYNTVEVGNKVMWIFINESDDTVSAPAPAKVA 141
Db 606 -----VSLNGTSGSLQFTTK-RTNK-YAITNKFEGTLNISGK--VNI 643
QY 142 APAAPAKQOGC-----RTVQVRSRIQTLYPGKTTAAAPFTESVVVSAPESPAKQAAA 196
Db 644 SWVLPRNMSGYDKFGRTYWNLTSLNVSSEGFNLTDISRGSDSAGTLTPYN----- 696
QY 197 ASAKQQTAAAPAKQQTAAAPAKQQTNTIDFRKDGKKNAGIIEALALGFA----- 248
Db 697 -----LNGISFNKDTTFNVERNA-----RVNFDIKAP---IGINKYSSLNYASFNGNISV 743
QY 249 ---GQPDISQQDHIIIVTLKNNHTLPTTLQRS--LDVAD-----FKTPEVCKVTLKRLNNDT 298
Db 744 SGGGSVDFT-----LLASSNNVQTGCVINSKYFNVSITGSSLRFFKTSGGTKTGFSIEKDL 798

```

|      |  |      |
|------|--|------|
| 799  | TL-NATGGNITLLQVEGTGMIKGIVAKKNITFEGGNI-----TFGSKAVTEIEGNVT        | 852  |
| 357  | IRTILQILAKESGMINIVASDVNGKMTLSLKQVPWDCALDLVMAQANLDMRQOGNIWNIA     | 416  |
| 853  | IN-----NNANVTLLIGSDFDNHQRPILIKK-----DVIINSGL--TAGGNIVNIA         | 896  |
| 417  | PRDELLAKDAFLQABEKDIA--DLGALYSONFQLKYKNVEEFRLILDLNADTTGNRNTLV     | 475  |
| 897  | --GNLTVESNANFKAITNTFNVGGLF-----DNKXGNSNIS                        | 931  |
| 476  | SGRGSVL--IDPATNLTIVDTRESVIEKRRKIID-----ELDVPAQ                   | 514  |
| 932  | AKGGAARFYDINSKNLSITNNSST--YRTILISGNIITKNGLNITNDESGDTMQLGGD       | 988  |
| 515  | QVMIEARIVEAAD--GFSRDLGVKFGATGKKKLKNDTSAPFGWGVNSGFGDDKKGWAETKI    | 572  |
| 989  | VSQKEGNLTISSDKINITKQITIKAGVDGENSDSDATN-----                      | 1026 |
| 573  | NLPITAAANSISLVRAIS--SGALNLELSASE-----SLSKTKTL--ANPRVLTQNR-KEAK   | 624  |
| 1027 | NANLTITKTKELKTQDLNITSGFNKAEITAKGDSDLTITGNNSADGTVNAKVTITNOVKDSK   | 1086 |
| 625  | IES-GYEIPPTVTSIANGSGSSTNTE-LKKAVALGLTV--TPNITPDGQIIMTVTKINKGSPAQ | 681  |
| 1087 | ISADGHKVTLH--SKVETSCGNNNTEDSSDNNAGLITDAKNVTVNNNITS-----HKAVSIS   | 1141 |
| 682  | CASGNQTLICISTKNL---NTQAMVENGGLTIVGGIYEEDNGNLT                    | 724  |



|  |      |   |      |
|--|------|---|------|
| Db   | 1142 | ATSGEITTKTGTINATTGNVEITAGTGS--ILGGIESSGSGSVTLT                  | 1185 |
| <p>RESULT 25</p> <p>US-08-719-641-2</p> <p>Sequence 2, Application US/08719641</p> <p>Patent No. 6218141</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Barenkamp, Stephen J</p> <p>TITLE OF INVENTION: High Molecular Weight Surface Proteins</p> <p>TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus</p> <p>NUMBER OF SEQUENCES: 10</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: Shoemaker and Mattare, Ltd.</p> <p>STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza</p> <p>CITY: Arlington</p> <p>STATE: Virginia</p> <p>COUNTRY: U.S.A.</p> <p>ZIP: 22202-0286</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Floppy disk</p> <p>COMPUTER: IBM PC compatible</p> <p>OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>SOFTWARE: Patent in Release #1.0, Version #1.30</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/08/719,641</p> <p>FILING DATE:</p> <p>CLASSIFICATION: 530</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/302,832</p> <p>FILING DATE: 16-SEP-1994</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US PCT/US93/02166</p> <p>FILING DATE: 16-MAR-1993</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: GB 9205704.1</p> <p>FILING DATE: 16-MAR-1992</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: Berkstresser, Jerry W</p> <p>REGISTRATION NUMBER: 22,651</p> <p>REFERENCE/DOCKET NUMBER: 1038-625</p> <p>TELECOMMUNICATION INFORMATION:</p> <p>TELEPHONE: (703) 415-0810</p> <p>TELEFAX: (703) 415-0813</p> <p>INFORMATION FOR SEQ ID NO: 2:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 1536 amino acids</p> <p>TYPE: amino acid</p> <p>STRANDEDNESS: single</p> <p>TOPOLOGY: linear</p> <p>MOLECULE TYPE:</p> <p>US-08-719-641-2</p> <p>Query Match 4.0%; Score 155.5; DB 3; Length 1536;</p> <p>Best Local Similarity 19.1%; Pred. No. 0.0017;</p> <p>Matches 146; Conservative 123; Mismatches 284; Indels 213; Gaps 41;</p> |      |   |      |
| Qy   | 25   | GNITDIKVSLLPNKQIKVKVSFDKVIWNPFTGVTSSPARIALDPBOTGISMDQQVLEYAD    | 84   |
| Db   | 567  | GNI-----NITAKQDIAEKGNSQVITGQTIISG-----NQKGFERN-----             | 605  |
| Qy   | 85   | PLLSKISAAQNSSRARLVNLNKPGE---QYNTVEVRGNKVMIFINESDDTVSAPARPAVKA   | 141  |
| Db   | 606  | -----VSLNGTSGSLQFTTK-RTNK-YAINTKFEGLTINISGK--VNI                | 643  |
| Qy   | 142  | APAPAKQOQC-----RTVYQVRSIRIOTLYPGKTTAAAPFTESVVSVAPEFSAKQCAA      | 196  |
| Db   | 644  | SMVLPKVESGYDKFKRTYWNLTSLNVSSEGFNLTIISRGSDSAGTLITQPIN-----       | 696  |
| Qy   | 197  | ASAKQOOTAAPAKQOOTAAPAKQOOTAAPAKQOOTAAPAKQOOTAAPAKQOOTAAPAKQOOTA | 248  |

|  |      |   |      |
|--|------|---|------|
| Db   | 1142 | ATSGEITTKTGTINATTGNVEITTAQTGS--ILGGIESSGSGSVTLT                 | 1185 |
| <p>RESULT 25</p> <p>US-08-719-641-2</p> <p>Sequence 2, Application US/08719641</p> <p>Patent No. 6218141</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Barenkamp, Stephen J</p> <p>TITLE OF INVENTION: High Molecular Weight Surface Proteins</p> <p>TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus</p> <p>NUMBER OF SEQUENCES: 10</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: Shoemaker and Mattare, Ltd.</p> <p>STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza</p> <p>CITY: Arlington</p> <p>STATE: Virginia</p> <p>COUNTRY: U.S.A.</p> <p>ZIP: 22202-0286</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Floppy disk</p> <p>COMPUTER: IBM PC compatible</p> <p>OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>SOFTWARE: Patent in Release #1.0, Version #1.30</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/08/719,641</p> <p>FILING DATE:</p> <p>CLASSIFICATION: 530</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/302,832</p> <p>FILING DATE: 16-SEP-1994</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US PCT/US93/02166</p> <p>FILING DATE: 16-MAR-1993</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: GB 9205704.1</p> <p>FILING DATE: 16-MAR-1992</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: Berkstresser, Jerry W</p> <p>REGISTRATION NUMBER: 22,651</p> <p>REFERENCE/DOCKET NUMBER: 1038-625</p> <p>TELECOMMUNICATION INFORMATION:</p> <p>TELEPHONE: (703) 415-0810</p> <p>TELEFAX: (703) 415-0813</p> <p>INFORMATION FOR SEQ ID NO: 2:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 1536 amino acids</p> <p>TYPE: amino acid</p> <p>STRANDEDNESS: single</p> <p>TOPOLOGY: linear</p> <p>MOLECULE TYPE:</p> <p>US-08-719-641-2</p> <p>Query Match 4.0%; Score 155.5; DB 3; Length 1536;</p> <p>Best Local Similarity 19.1%; Pred. No. 0.0017;</p> <p>Matches 146; Conservative 123; Mismatches 284; Indels 213; Gaps 41;</p> |      |   |      |
| Qy   | 25   | GNITDIKVSLLPNKQIKVKVSFDPKTEIWNPTGFTVTSPPARIALDPBOTGISMDQQVLEYAD | 84   |
| Db   | 567  | GNI-----NITAKQDIAIEKGSNQVITQGTITSG-----NQKGFERN-----            | 605  |
| Qy   | 85   | PLLSKISAAQNSSRRARLVNLNKPGE---QYNTVEVRGNKVMIFINESDDTVSAPARPAVKA  | 141  |
| Db   | 606  | -----VSLNGTSGSLQFTTK-RTNK-YAINTNKFEGTLNISKG--VNI                | 643  |
| Qy   | 142  | APAPAKQKQC-----RTVYQVRSIRIOTLYPGKTTAAAPFTESVVSAPFSPAKQCAA       | 196  |
| Db   | 644  | SMVLPKVESGYDKFKRTYWNLTSLNVSSEGFNLTIISRGSDSAGTLITQPIN-----       | 696  |
| Qy   | 197  | ASAKQQTAAAPAKQQTAAAPAKQNTNIDFRKDGNGNAGIIEAALGFA-----            | 248  |

142 APAAPAKQGC-----RTYQVRSIRIQLYPGKTTAAAPTESVVSYPSPAKQQA 196  
Db 644 SWMLPKNESGYDKFGKTYWNLTSLNBSGSEFNLTIDSRGSDSAGITQPIN----- 696  
QY 197 ASAKQCTAAPAKQCTAAPAKQQAAPAKQTNIDFRKDGKNAGIIEAALGFA----- 248  
Db 697 -----LNGISFNKDTTFNVERNA-----RVNFDIKAP-----IGINKYSILNVSFNGNISV 743  
QY 249 ---GQPDISQGDHIIIVTLKHNHTLPTTLORS--LOVAD-----FXTPPQKVTKLKLNNDT 298  
Db 744 SGGGSVDFT-----LIASSNVQTPGVVNSKYFNVSCTGSSLRFTSGSTGTGFSIEKDL 798  
QY 299 QLIITTAGHNELVNGAAGCYFTFQVLPKKQ-NLESQGVNNAKPTFTTGKISLDFQ-DVE 356  
Db 799 TL-NATGGNITLLQVSTGDMICKGIVAKKNITFEQGN-----TFGSRKAVTEIEGNVT 852  
QY 357 IRTILOLAKESGMNIVASDVNGKMTLSLKVDHPDOALDLVMQARNILDMRCQGNIVNIA 416  
Db 853 IN-----NNANVTLLIGSDFDNHQKPLTIKK-----DVIISGNL--TAGGNIVNIA 896  
QY 417 PRDELLAKDAFLQAEDKA-DGALYSONFQLYKXNWEFRSILRLONADTIGNRNLV 475  
Db 897 --GNLVESNANFAKAITNFTFNVGGLF-----DNKGSNISI 931  
QY 476 SGRGSVL--IDPATNLTIVTDRSVTEKPKKLID-----ELDVPAQ 514  
Db 932 AKGGARFKDIDNSKNLSITNNSST---YRTIIISGNITKNKGLNITNEGSTDEWQIGD 988  
QY 515 QVMTEARIVEAAD--GFSRDLGVKFGATGKKLKNDTSAFGMVNSGFGGDDKWAETKI 572  
Db 989 VSQEGNLTISSDKINITKQITIKAGVDGENSDSDATN----- 1026  
QY 573 NLPITAAANSISLVRAIS-SGALNLELSASE-----SLSKYKTL--ANPRVLTONR-KEAK 624  
Db 1027 NANLTIKTKELKLTQDLNISGFNKAEBITAKDGSDLTIGNTNSADGNTAKKVTFNQVYKSK 1086  
QY 625 IES-GYEIPFTVTSIANGSGSSTNTE-LKKAVALGLTV-TPNITPDGOIIMTVKINKDSPAQ 681  
Db 1087 ISADGHKVTLH-SKVETSGSNNITEDSSDNNAGLTIDAKNVTNNNITS-----HKAVSIS 1141  
QY 682 CASQNTQILCISTKNL---NQAMVENGGLIVGGIYEEDNGMNTL 724  
Db 1142 ATSGEITTKGTGINATTGNVEITAGTGS--ILGGIESSGSVTLT 1185  
  
RESULT 27  
US-09-252-991A-32469  
; Sequence 32469, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/352.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 32469  
; LENGTH: 990  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32469  
  
Query Match 4.0%; Score 152.5; DB 4; Length 990;  
Best Local Similarity 20.7%; Pred. No. 0.0015;  
Matches 150; Conservative 106; Mismatches 248; Indels 221; Gaps 36;  
OV 204 AAPAKQCTAAPAKQQAAPAKQTNIDFRKDGKNAGIIEAALGFA----- 249

```

2  AAPEWRVTQWRPAQPAQRRHRRSGRRIGSGGQLEQPGTPPEREERSAGQPGRCSLQR 61
250  ----OPDISQQDHIIIVTLKHHTLPTTLQRSLLD----- 278
62  RDLVAFPGCVGEPGGGVVTDAGAVLRS- --SLDNSQGGIVSAKGAABIRTGSLNNSOK 118
279  ----VADFPTPVQ--KVTK- ----RLNNDTQLIITTAGNWE 309
119  GGIGSGAGLAVDLVDNSNGRITAKGAIDANLKGDLQOGSGRLVSDTAIALDURGG-E 177
310  LVNKS----AARGVYTFQVLPRKQNLSESGVNNAPKTF-----GRKISLDFQDV 355
178  LVNKAQGLIATPGALLRLQGVVDN--SGGEISDSRSPFLAATALSNRGRVIGSLSLT 236
356  EIRTIQ-----ILAKESGMNIVASDSVNGKMTLSLKDVFWQALDLVWQARNLDMRQOG 410
237  RIAQALDNSLQGVLSAGGLDVAALVFNHSGIVASKG-----DTRIGVNRLE-NEAG 288
411  NVNVIAPRELLAKD---KAFICAEDI-ADIGALYSONFOLKYKNVEEFRSILRLDNA 465
289  RVWSESALD-LTAKQVSSAKGRTAAKGDLCVTVGTLEQQGGELASQGT-----LTLD-A 340
466  DTTGNRN-TLVSGRGSVL-----IDPATNTLIYDTRSVIEKPKLID-----BL 509
341  DSLNRRNGLSVADGGVTAEARQIDNRGGE--ISSVAKVALAVREQLDNRGGKVIQDSEL 398
510  DVPAQQWMIAR-IVEAADGFSRDLGVKFGATGKKLXNDTSFAFGWGVNSFGGDDKWGA 568
399  SLTVQRLNQAKVLAGRDLHLHD-----GAEILINGD-----GG--LLSS 436
569  ETKINLPITAAANGISLVRAISSGALNLELSASELSKTKTLANPRVLTONRKEAKIESG 628
437  QELVDVTLISGALDNOGSGALVSEGSLLVKADQVNNQAGTSPSSAGSLLVT--SRGELNNOGQ 495
629  YEIPTVTSIANGSGSTN--TELKXAVLGLVTPNITPDGQIINTVKINDKSPAQCASNQ 687
496  -----RLVTDAGATLNTSGTFDNRAGL-----VSAGGAVARTGAALNNSQKSGISGGNT 543
688  TILCIS-----TKXLTQAMVENGGLTIVG--GIYEEEDNGNTL----- 723
544  GTVLVAGLVNDRGREGRISTQGTLDANLKGLLHGGSSLVGERGVTLDLNGGTLDNHDLGL 603
724  --TKVPLL-----GDI-----PVIGNLFKTRGKTKDRRELLIFITPRIMGTA 763
604  VSTPGALLRLQLGMVDNSVGGEISSDRAFTLAAANTLNNQGRLLISSEAL---TLRIAKTL 660
764  GNSLR 768
661  DNSLK 665

RESULT 28
US-08-728-470-9
; Sequence 9, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Martare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

Query Match      4.0%; Score 152.5; DB 4; Length 990;
Best Local Similarity 20.7%; pred. No. 0.0015;
Matches 150; Conservative 106; Mismatches 248; Indels 221; Gaps 36;
Ov 204 AAPAKOOTAPAKOAAAAPAKOTNIDFRKDKGNAGIEL-----AALCFAG----- 249

```

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/728,470  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-633  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1338 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-728-470-9

Query Match 3.9%; Score 150; DB 2; Length 1338;  
Best Local Similarity 21.6%; Pred.No. 0.0039;  
Matches 169; Conservative 105; Mismatches 312; Indels 198; Gaps 37;

QY 103 LNLNKGQVNTVRGNKVMIFINESDDTVSAPAPAVKAA-----PAAPAKQQCGRT 154  
DB 412 LNVTSGRFNLSI-----DSTGSGTGFPSIRNAELNGITTFNKATFNIAQGSTA 459  
QY 155 VTQVRSRIQTLYPKGTAAAPFTESVSAPSAPKAQAASAKAQT----- 203  
DB 460 NFSIKA---SIMPTKSNANYALFNEDISVGSGSVNFKLNASNNIQTPGVIIKSQNFN 515  
QY 204 -----AAPAKQOTAPAKQAAAAPAKQTNIDFRK-DGNAGAIIELAALGFAGQPDI 253  
DB 516 VSGGSTLNLKAEGSTETAFSIENDLNLTATGNTITIROVEGTDNRVNK---GVAACKNI 571  
QY 254 SQOQHII-----VTLLKNHLPPTLQRSLDVADFKTPVKVLTKRLNNDTLQ 300  
DB 572 TFGKGNITFSQKATTEIKGNVTINKNTNATL--RGANFAENKSP-L-NIAGNVNNGNLT 628  
QY 301 ----IITTAGNELVNKSA---APGYFFOVLPKKONLESQGVNNAKPTFTGRKI----- 348  
DB 629 TAGSIINIAGNL-TYSKGANLQAITNYTFNV---AGSFENNAGSNISIARGAKFKDINN 684  
QY 349 --SLDP---ODVEIRTI-----LOILAKES-----GMNI-----VASDSV 378  
DB 685 TSLSNITNTSDTYRTIIGKNISNKSGLDIIDKSKDAEIQIGNISQKEGNLTISSDKV 744  
QY 379 NGKWTLSLK-----DYWPQDALDVMAARNLMQRQGNIVNIAPRDELLAKD----- 425  
DB 745 NITNQITIKAGVEGGSRSSSEAEANLTIQTELKLAGDLNISGF-NKAEITAKNGSDLT 803  
QY 426 -----KAPLQAEKD--IADLGALYSQNOLKYKN-----VEEFRSILRLDNAD 466  
DB 804 IGNASGNADAQKVTEKVDKSKISTDGHNVTLNSEVKTNSGNSNAGNDNSTGLTISAKO 863  
QY 467 TTGNRNLTVSGRGSVLIDPATWTLIVDFRSVIEFKRLIDELDYPAQOVMEARIIVEEA 526  
DB 864 VTVMNN--VTSHKTNISAAAGNVTTKEGTTI----NATTGSVEVTAQNGTKIGNIT--- 914  
QY 527 DGFSDRLGVKFGATGKKLKNDITSFSGWGVNSGFG-----GDGRKGAEE-TKINLPIT 577  
DB 915 ---SQNVTV---TATENLVTTENAV--INATSGVVIWISTXGTGDKIGSISTSGNVNIT 964

```
Db 412 LNVTSKGFNLIS-----DSTGSGTGSPSIRNAELNGITFNKATFNIAQGSTA 459
QY 155 VYQVRSIRIOTLYPGKTTAAAPFTESVVSAPSPAKQQAASAKQOT-----203
Db 460 NFSIKA-----SIMPFSKNANYALFNEDISVSGGVSFNKLNASSNNIQTGVIKSNQFN 515
QY 204 -----AAPAKQOTAAAPAKQQAAPAKQNTIDFRK-DGKNAGIIEALALGAFAGQPD 253
Db 516 VSGGSTLNLKAEGSTETAFSIENDLNLNATGNTITIQVEGTDSDRVNK-----GVAAKKNI 571
QY 254 SQQHDHII-----VTLKNHTLPTTLQSLDADFKTPVQKVTLKRNLNDTQL 300
Db 572 TPKGNITFGSQKATTEIKGNVTINKNTNATL--RGANFAENKSP--NIAGNVINNGNLT 628
QY 301 -----IITAGNWLNVKSA---APGYTFQVLPKQNLSEGGVNNAPKTTGRIK-----348
Db 629 TAGSIINIAGNL-TVSKGANLQAITNYTFNV---AGSFDNNGASNISARGGAKFKDINN 684
QY 349 --SLDF--QDVEIRTI-----LOILAKES-----GMNI-----VASDSV 378
Db 685 TSSLNITNSDTTYRTTIKGNISKSGDLNIDKSDAEIOIGNISQKEGNLTISSDKV 744
QY 379 NGKMTLSLK-----DVPWDQALDVMQARNLDMROQGNIVNIAPRDELLAKD-----425
Db 745 NITNQITIKAGVEGGRSDSSEAEANLTIQTKELKLAGDLNISGF--NKAEITAKNGSDLT 803
QY 426 -----KAFLOAKED--IADLGALYSONFOLKYKN-----VEEFRSILRLDNAD 466
Db 804 IGNASGGNADAKKVTDFKVKDSKISTDGHNVTLNSEVKTNSGSSNAGNDNSTGLTISAKD 863
QY 467 TTGNRNTLVSGRGSVLIDPATNTLIVDTRSVIEKFRKLIDELVPAQOVMIARIVEAA 526
Db 864 VTWNN--VTSHKTINISAAAGNVTTKEGTII---NATTSQVEVTAQNGTIKGNIT---914
QY 527 DGFSRDLGVKFGATGKKLKNDSAFGWSGFG-----GDDKWGAE--TKINLPIT 577
Db 915 ---SQNVTV---TATENLVTTENAV---INATSGTVNISTKTDIGKGIESTSGNVNIT 964
QY 578 AAANSISLVR-----AISSGALNIELSASBSLTKTLANPRVLTO--NRKEAKIESGY 629
Db 965 ASGNTLKVSNIQTQDVTVTADAGALTTTAGTISATTNAGNANITTKTGDINGKVESSSGV 1024
QY 630 EIPFTVTSANGSSNTTELKAVLG-LVVTENITPDGQILMTVKINKDSPAQASGNQT 688
Db 1025 TLVATCATLAVNIGNVTITITADSGKLTSTVGSTINGNTSVTSSQSGDIEGTISGNTV 1084
QY 689 ILCISTKNL--NQAMVE--NGG-----TLVGGIYBEDNGNTLTKVPLLGDPV 734
Db 1085 NVTASTGDLTIGNSAKVEAKNGAATUTAESKLTITQTGSSITSSNGQITLTAK---DSSI 1141
QY 735 IGNU 738
Db 1142 AGNI 1145

RESULT 30
US-08-617-697-9
; Sequence 9, Application US/08617697
; Patent No. 5977336
; GENERAL INVENTOR:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1599 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-9

Query Match 3.9%; Score 150; DB 2; Length 1599;
Best Local Similarity 21.6%; Pred. No. 0.0051;
Matches 169; Conservative 105; Mismatches 312; Indels 198; Gaps 37;

QY 103 LNLKPGQYNTVVRGNKVIWFINESDDTVSAPARPAVKAA-----PAAPAKQOGCRT 154
Db 672 LNVTSKGFNLIS-----DSTGSGTGSPSIRNAELNGITFNKATFNIAQGSTA 719
QY 155 VYQVRSIRIOTLYPGKTTAAAPFTESVVSAPSPAKQQAASAKQOT-----203
Db 720 NFSIKA-----SIMPFSKNANYALFNEDISVSGGVSFNKLNASSNNIQTGVIKSNQFN 775
QY 204 -----AAPAKQOTAAAPAKQQAAPAKQNTIDFRK-DGKNAGIIEALALGAFAGQPD 253
Db 776 VSGGSTLNLKAEGSTETAFSIENDLNLNATGNTITIQVEGTDSDRVNK-----GVAAKKNI 831
QY 254 SQQHDHII-----VTLKNHTLPTTLQSLDADFKTPVQKVTLKRNLNDTQL 300
Db 832 TPKGNITFGSQKATTEIKGNVTINKNTNATL--RGANFAENKSP--NIAGNVINNGNLT 888
QY 301 -----IITAGNWLNVKSA---APGYTFQVLPKQNLSEGGVNNAPKTTGRIK-----348
Db 889 TAGSIINIAGNL-TVSKGANLQAITNYTFNV---AGSFDNNGASNISARGGAKFKDINN 944
QY 349 --SLDF--QDVEIRTI-----LOILAKES-----GMNI-----VASDSV 378
Db 945 TSSLNITNSDTTYRTTIKGNISKSGDLNIDKSDAEIOIGNISQKEGNLTISSDKV 1004
QY 379 NGKMTLSLK-----DVPWDQALDVMQARNLDMROQGNIVNIAPRDELLAKD-----425
Db 1005 NITNQITIKAGVEGGRSDSSEAEANLTIQTKELKLAGDLNISGF--NKAEITAKNGSDLT 1063
QY 426 -----KAFLOAKED--IADLGALYSONFOLKYKN-----VEEFRSILRLDNAD 466
Db 1064 IGNASGGNADAKKVTDFKVKDSKISTDGHNVTLNSEVKTNSGSSNAGNDNSTGLTISAKD 1123
QY 467 TTGNRNTLVSGRGSVLIDPATNTLIVDTRSVIEKFRKLIDELVPAQOVMIARIVEAA 526
Db 1124 VTWNN--VTSHKTINISAAAGNVTTKEGTII---NATTSQVEVTAQNGTIKGNIT---1174
QY 527 DGFSRDLGVKFGATGKKLKNDSAFGWSGFG-----GDDKWGAE--TKINLPIT 577
Db 1175 ---SQNVTV---TATENLVTTENAV---INATSGTVNISTKTDIGKGIESTSGNVNIT 1224
```

QY 578 AAANSISLVR-----AISSGALNELSASBSLSKTKTLANPRVLQ--NRKEAKIESGY 629  
Db 1225 ASGNTLKVSNTIGQDVTYADAGALTITAGSTISATTGNANITTKTGDIKNGKVESSSGSV 1284  
QY 630 EIPFTVTSIANGSGSTNELKXAVLG-ITVTENITPDGQIIMTVKINKXDSQAQASGNQT 688  
Db 1285 TLVATGATLAVGNISGNTVTITADSGKLTSTVGSTINGTNSVTSSQSGDIEGTISGNTV 1344  
QY 689 ILICISTKXL--NTQAMVE--NGG-----TLIVGGIYEEDNGNTLTVKVPLLGDIPIV 734  
Db 1345 NVTASTGDLTIGNSAKVEAKGAATLTABSGKLTITQTGSSITSSNGQTTLTAK---DSSI 1401  
QY 735 IGNL 738  
Db 1402 AGNI 1405

RESULT 31  
US-09-120-663-2  
; Sequence 2, Application US/09120663  
; Patent No. 6228644  
; GENERAL INFORMATION:  
; APPLICANT: Bogdanove, Adam J.  
; APPLICANT: Kim, Jihyun Francis  
; APPLICANT: Wei, Zhong-Min  
; APPLICANT: Beer, Steven V.  
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM  
; TITLE OF INVENTION: ERWINIA AMYLOVORA, ITS USE, AND ENCODING GENE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: P.O. Box 1051, Clinton Square  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patencin release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/120,663  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/055,105  
; FILING DATE: 06-AUG-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/1661  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (716) 263-1304  
; TELEFAX: (716) 263-1600  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1838 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-120-663-2

Query Match 3.8%; Score 147.5; DB 3; Length 1838;  
Best Local Similarity 18.9%; Pred. No. 0.01;  
Matches 146; Conservative 106; Mismatches 330; Indels 189; Gaps 29;  
QY 20 QTAS----AGNITDKVSSLPNKQIKVKSFPDKIIVN-----PTGFTVTSIPARIALDPE 69  
Db 814 QTSREGISGELKDIHD---HKQNLVATLHEGVFHPQREAWNGAESWTKLALPQS 870  
QY 70 QTGISMDQOVLEYADPLLSKISAAQNSRARLVNLNKPQYINTEVRGNKVWIFINESDD 129

Db 871 E-----SKLKSLSMSHEHKPIATFEDGSOHQLKAGG-----W-----H 903  
QY 130 TVSAPARPAVKAAPAKQCGCTVYVRSIRIOTLYPGKTTAAAPFTESVVS----- 183  
Db 904 AYAAPER-----GPLA-VGTSGSQTFFN-----SLMQGVKGVIPGSGLTIVKLSAQTOGMT 953  
QY 184 -----VSAPFSPAKQQAASAKQOATAAPAKQQTAA-----PAKQQAAPAKQ 225  
Db 954 GAERKVSSEFSEIRIAYAFNPWTSTPRPKIKNAYATQHWQWQREGKLPLEYOGALIKO 1013  
QY 226 TNIDFRKDGKNAGIIELAAALGFAGQPDISOQHDHIIIVTLKXHTLPTTLQRLSDVADPKTP 285  
Db 1014 --LDAHNVREN-----APQPDLSQKLETLDLGHEGAELLNDMKFRD----- 1053  
QY 286 VQKVTLRNLNDIQLIITTAGNWLKNSAAGVFTFQVLPPKQNLSEGGVNNAPKFTTG 345  
Db 1054 -----ELQSATRSVTVLGQHGVLKSNSEINSEPKSPGKALVGSFVNR-----SG 1101  
QY 346 RKISLDFQDVEIRTIQLAK-ESGMNIVASDSYN-----GKMTLSLKDPWQDQALDVM 399  
Db 1102 QDLSKSLQAAVHATPPSAESKLSQMLGHFVSAGVDMSHQKEIPLGRQRPNDKTA-LTK 1160  
QY 400 QARNLDMRQQGNIVNIAPRELLAKDKAFLOAEKDI-----ADLG 439  
Db 1161 SRLILDTVTIGELHELADKAKLVSDHKPDADQIKOLEQQQFDTLREKRYESNPVKHYTDMG 1220  
QY 440 ALYQNFQKLYKNVEEERSILRLDNADTTGNRNLTVSGRGSV-LIDPATNTLIIVTDRSV 498  
Db 1221 PTHKALEANYDAVKAFINAFKKEHGVNLTRIVLESQSAELAKLKNLTLSDSGES 1280  
QY 499 IEKPRKL---IDELDVPAQQVNIARIVEAAD-----GFSR---DLQVKEG----- 538  
Db 1281 MSFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAYNLSFSRTSGGLNVSEGRDGV 1340  
QY 539 -----ATGKKLKNDTSAFGWYNSGFGGDDKW-GAETKINLPITAAANSISLVAISS 591  
Db 1341 SGNIMVATG-----HCVNPMYMTGKTSAGNASDMSAKHKISPLRIGAAVSGTLQTLQ 1395  
QY 592 GALSLELSASE-----SLSKTTLANPRVLTQNRKEAKIESGYEIPFTVTSIAN----- 640  
Db 1396 NSLKEFLTEDELPGFIHGLTHGLTTPAELLQKGIHQWKQSKLTFSDVTSANLDRAGI 1455  
QY 641 ----GGSTNTTELKXAVLGTLVTPNI-----TPDQIIMTVKINKDSP 679  
Db 1456 NLNEDGSKPNGVTVARVSAGLSANLAAGSRERSTTSCQFGSTTSASNNRP 1506

RESULT 32  
US-09-328-352-4412  
; Sequence 4412, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4412  
; LENGTH: 734  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
; US-09-328-352-4412

Query Match 3.8%; Score 146.5; DB 4; Length 734;  
Best Local Similarity 19.3%; Pred. No. 0.0029;  
Matches 141; Conservative 109; Mismatches 286; Indels 193; Gaps 32;  
QY 32 VSSLPNKQIKVKSDEKIVNPTG---FVTSSPARIALDPEOTGISMDQOVLEYADPLL 87  
Db 57 VIHLKNGETIVLENFNEATNTTEHSLVFPTEQGFVEAQFDAQKVIDYRGLNHVTDLA 116

```

Query Match      3.8%; Score 146.5; DB 4; Length 1180;
Best Local Similarity 20.8%; Pred. No. 0.0061;
Matches 101; Conservative 80; Mismatches 182; Indels 122; Gaps 27;

QY 88 -----SKISAAQNGSRRLVLNLNKPQYQYNTVRGNKVMWIFINESDDTVSAPARPAKVA 141
Db 117 YTSSTSPSAATMAVDNDPSFGMGVNLKAGLAVLAAEGLYLMAF--DKDDKDDSPSTPOL-I 173
QY 142 APAAP-----AKQGCRTVYQVRIRIQTLTPGKTTAAAPFTESVVSAP-- 187
Db 174 APAAPTATLADDTVTGKTEANAKIY-IKDAAGNTVAGSVADASGNYT---IKLDKPLV 229
QY 188 ----FSPAQQAAASAKQOTAPAKQCTAAAPAKQO-----AAPAKQCTNIDF 230
Db 230 NGDKLVIAQDAAGNSKVTVVTGKDTIAPDVPQALSDGSLLTGKAEANAKITVYD- 298
QY 231 RKDGKAGIIEALAGFAGOPDISQHDHIIIVTLKNHLPITLQOR-----SLVADFKTP 285
Db 289 -ATGKVLGTV-----FANKOG-----YLSLK-LTPPLTSEAGKVVAEDAAAGNSE 332
QY 286 VQKVTLLKRLNNDTLQIITAGNWLNVKSAAPGYFTFQVLPKPKQNLSEGGVNNAPKFTTG 345
Db 333 EVKI-----IAGKDTIPPASPFFVENKEGVIHG-----KTEAN 366
QY 346 RKISLDFQDVEIRTLQILAKESGMNIVASDSVN--GRWTLSLKDVDPQALDLVMQARN 403
Db 367 AKQVI-----KDRGKVLGSGTADAQGEFQITL-----SPALKEAQKGTV 406
QY 404 LDMROCGNI---VNIAPRDELLAKAFLOAEKDIADLGALYSQNFOLKYKNVEEFSIL 460
Db 407 VVEDAAGNVKPEVEITPGFDSIAPDKPTVQINTDGTSTVGTAEANAKIEIK----- 457
QY 461 RLNDADTT---GNRLTVSGRGSVLIDPATNLTIVTDTT---SVTEKPRKLIDELDV- 511
Db 458 -----DTCKVLSGSDADANGFTSISPA-----LTENKIASVADNAGNKSEVVDIV 507
QY 512 -----PAQQWIEARIVEAADGRSRLGVKFGATGKKLKNKNTSAPFGVNVSGFGGDD 564
Db 508 GTRDTTPPAKPII-----NSVDDDDVGAVKGAITAGSETDD 542
QY 565 ---KMGAEKINLPITAAANSISL-VRAISSGALNLELSASELSKCKTLANPRVLQNR 620
Db 543 ARPKLTSGEANATLIYDNGVAIGVTVTSR-----SWSTFDKOLALGXHTITLTQT 597
QY 621 KEAKIESGVEIPTVTSIANGSSNTLKKAVLGLTVPNITPDGQIIMTVKINKDSPA 680
Db 598 DAAGLTSEASSPFTFYVAPKAASL-SETSDVILS-TEGPSLA-DSVGLHLLTKVAQNTTT 654
QY 681 QCASGNQTI 689
Db 655 ETNNPKSV 663

RESULT 33
US-09-206-942-65
; Sequence 65, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 65
; LENGTH: 1180
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-65

```

```

Query Match      3.8%; Score 146.5; DB 4; Length 1180;
Best Local Similarity 20.8%; Pred. No. 0.0061;
Matches 101; Conservative 80; Mismatches 182; Indels 122; Gaps 27;

QY 294 LNNDTQLIITAGNWLNVKSAAPGYFTFQVLPKPKQNLSEGGVNNAPKFTTGRKISLDFQ 353
Db 357 INNDLTL-NATGNGNISLQVEGIDMGEGVAAK-NI-----TTGGNITFGSK 404
QY 354 DVEIRTIQI-----LAKESGMNIVASDSVNGKNTLSLKDVDPQALDLVMQARNLDMRQ 408
Db 405 ----KAITEIKGNVTINENTNATLIGSDFNDHKKPLNKG-----DVNVRG---NLTA 450
QY 409 QGNIVNIAPRDELLAKAFLOAEKDIA-DLGALY-----SONFOL-----KYKNVEEFS 458
Db 451 GGNVINIG--GNLTVENGANLKAITNFTFVGGFLPNKGNISISIAARGGAKFKDINTSS 508
QY 459 ILRLDNADTGNRLTVSGRGSVLIDPATNLTIVTDTSTRSVIEKPRKILDELDPVPAQVMI 518
Db 509 LNIITNSDTT--YRTIIEGN---ITNKAGDLNIDNKGNA-----EIQIGGNISQK 554
QY 519 EARIVEAAD--GFSRDLGVKFGATGKKLKNKNTSAPFGVNVSGFGGDKWGAETKINLPI 576
Db 555 EGNLTISSDKNITKQITKGVNGE---NSDSS-----TKSQANL 592
QY 577 TAAANSISL-----VRAISSGALNLELSASELSKCKTLANPRVLQNRKEAKI 625
Db 593 TIKTKELKLTODLNIISGPNKAKIVAKDSSNLTIGNSDDSGNTSA---KTVTFNNVKDSKI 649
QY 626 ES-GYEIIFP---TVTSIAN---GGSSTNTELKAVLGLTVTPNITPDGQIIMTVKINK 676
Db 650 SADGHVTLNKSVKLTSDNDNTEGSDNNT-----GLTIT---AKDEVNNTNITSHK 699
QY 677 DSPACASGNQTIILCISTKNL---NTQAMVENGTLVGGIYEEEDNGNTLTQVPLLEDIP 733
Db 700 TVNVSAANGGITTKTGTITNATAGNVEITAHTGS--IQGGIESKPGSVTIVAG---GDTL 754
QY 734 VIGNL 738
Db 755 AVGNI 759

RESULT 34
US-09-206-942-63
; Sequence 63, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER FILING DATE: 09/167,568
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 63
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-63

```

```

Query Match      3.8%; Score 146.5; DB 4; Length 1188;
Best Local Similarity 20.8%; Pred. No. 0.0061;
Matches 101; Conservative 80; Mismatches 182; Indels 122; Gaps 27;

QY 294 LNNDTQLIITAGNWLNVKSAAPGYFTFQVLPKPKQNLSEGGVNNAPKFTTGRKISLDFQ 353
Db 365 INNDLTL-NATGNGNISLQVEGIDMGEGVAAK-NI-----TTGGNITFGSK 412
QY 354 DVEIRTIQI-----LAKESGMNIVASDSVNGKNTLSLKDVDPQALDLVMQARNLDMRQ 408

```



|                       |                  |                  |             |              |
|-----------------------|------------------|------------------|-------------|--------------|
| Query Match           | 3.7%             | Score 143;       | DB 2;       | Length 1600; |
| Best Local Similarity | 21.4%            | Pred: No. 0.019; |             |              |
| Matches 132;          | Conservative 95; | Mismatches 234;  | Indels 156; | Gaps 33      |

  

|    |     |                 |                    |                 |                       |     |
|----|-----|-----------------|--------------------|-----------------|-----------------------|-----|
| QY | 224 | KOTNIDFRKGNKAGI | ---IELAALG----     | FAGQPD          | ISQDHHIIVLKNHTPTTLQRS | 276 |
|    |     |                 |                    |                 |                       |     |
|    |     |                 |                    |                 |                       |     |
| Db | 809 | KOTKDSFYNEYSKHA | INSHNLTLGGNVTLGGNS | SSITGNINIKAN--- | VTLQ---               | 863 |
|    |     |                 |                    |                 |                       |     |
| QY | 277 | LDAVAFKTPVQK-   | VTLKELNNDTLQIITTA  | -----GNWELNKSAA | PGYFTFQVLPKQ          | 329 |
|    |     |                 |                    |                 |                       |     |
| Db | 864 | ADTSNSNTGLKRTT  | TLGNIISVEGNSLITGAN | INVLGNIAEDSTPKG | -----                 | 912 |
|    |     |                 |                    |                 |                       |     |

```

RESULT 38
US-05-206-942-35
; Sequence 35, Application US/09206942
; Patent NO. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; TITLE OF INVENTION: Molecular Weight Proteins
; FILE REFERENCE: 1038-861 MIs:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-35

```



Db 125 KSAISIRGNVTKNLTGTGSAINIEKNLTVEGSAKFLANPNYSFNVSLFDNQGKS---181  
Qy 346 RKISLD-----FQDVEIRTIQILAKESGMNIVASDS-----VNGKWTLSLKDVPWDOAL 395  
Db 182 -NISIAGGAIFKDIENITGSLNITTK-----SDSNHHTIIGKNTNRKGD-----L 226  
Qy 396 DLVWQARNLDMROQNI-----VNIAPRDELLAKDKAFLOAEKDIADLGALY 442  
Db 227 NITNNGDTEIQIGGNISSQKVNITERITIKA-----GVNGDSDSNEAT 280  
Qy 443 SQNFQKYK-----NVIEFR--SILRLDNADTTGNRLTVSGRGSVLIDPATNTLIV 492  
Db 281 SANLTIKTELKLTNDLNSGFKAEITAKDNLNLTIGDN--SDAGN-- 325  
Qy 493 TDRSVIEKPKLID-ELDVPQOVMEARIVEAADGFRDLG-----VK 536  
Db 326 TDAKV--TFSNVKDSKISASCHNVTLNSKVETSGDSTDGNNNTGLTITAKVTVN 383  
Qy 537 FGATGKKL-----KNDTSFNGVNSGFG-----GDDKWAETKI-NLPITAAANS 582  
Db 384 NNTSHXTVNITASENVTTAGTTINATTGSEVETAKTGDIKGIESNGVNVITASGDT 443  
Qy 583 ISLVR-----AISSGALNLSASELSKTKTLANPRVLTONRK---EAKIESGYEI 631  
Db 444 LNVSNITQNVTVAAASGA---VTTTKGSTINATTGNANITTKTGEINGEVKSASG-NV 498  
Qy 632 PFT-----VTSIANGSGSSTNTELKAVLGLTVPNTIPDQGIIMTVKINK 676  
Db 499 NITASGNLNVSNITQNVTVTANSAGTITTE-----GSTI--NATTGDANITTOTGNI 550  
Qy 677 DSPAQACAGNOTILCISTKNTQAMVENGOTLIVGGIYEDNGNTLT 724  
Db 551 NGKVESSGSVT-----LIATGOTLAVGNI-----SGDTVT 581

RESULT 39  
US-09-206-942-37  
; Sequence 37, Application US/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; TITLE OF INVENTION: Molecular Weight Proteins  
; FILE REFERENCE: 1038-861 MIS:jb  
; CURRENT APPLICATION NUMBER: US/09/206,942  
; EARLIER FILING DATE: 1998-12-08  
; EARLIER FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 37  
; LENGTH: 1222  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-206-942-37

Query Match 3.7%; Score 141.5; DB 4; Length 1222;  
Best Local Similarity 21.4%; Pred. No. 0.016;  
Matches 135; Conservative 79; Mismatches 201; Indels 217; Gaps 33;  
Qy 239 IIELAALGFAGQPD-----SQQDHIIIVTLKNHT-----LPTTLQSL-- 277  
Db 328 IIESQNFASGSSSLKFKSEGSTHAFTIKNDLIINATGNIISLNOVAGIDSNLKSLIA 387  
Qy 278 -----DVADFKTPVQ-----KVTLKRLN--NDTQLI-----ITAGNW 308  
Db 388 NKNITFEGNITLAADKKPIEIKGNTVKEGANVTLRSANYGNDKSAISIRGNVTNKNL 447  
Qy 309 ELVNKS-----AAPGYFTFQVLPPKQNLSESGGVNAPKFTTGRKISLD-----FQDVE 356

Db 448 TVTGSAINIEKNLTVEGSAKFLANPNYSFNVSLFDNQGKS---NISIAGGAIFKDI 503  
Qy 357 IRTIQLILAKESGMNIVASDS-----VNGKWTLSLKDVPWDOALDLVWQARNLDMROQNI 411  
Db 504 NTGSLNITTK-----SDSNHHTIIGKNTNRKGD-----LNTNNGDNTIQTIGGN 549  
Qy 412 1-----VNIAPRDELLAKDKAFLOAEKDIADLGALYSONFQKYK----- 451  
Db 550 ISQKEGNTLISDDKVNITERITIKA-----GVNGDSDSNEATSANLTIKTELKLTND 603  
Qy 452 -NVIEFR--SILRLDNADTTGNRLTVSGRGSVLIDPATNTLIVTDRSVIEKPKLID- 507  
Db 604 LNISGFKAEITAKDNLNLTIGDN--SDAGN-- 646  
Qy 508 ELDPQOVMEARIVEAADGFRDLG-----VKFGATGKKKL-----KN 547  
Db 647 KISASCHNVTLNSKVETSGDSTDGNNNTGLTITAKVTVNNTSHXTVNITASEN 706  
Qy 548 DTSFNGVNSGFG-----GDDKWAETKI-NLPITAAANSISLVR-----AIS 590  
Db 707 VTTKAGTTINATTGSEVETAKTGDIKGIESNGVNVITASGDTLNVSNITQNVTVAAA 766  
Qy 591 SGALNLSASELSKTKTLANPRVLTONRK---EAKIESGVEIPFT----- 634  
Db 767 SGA-----VTTTKGSTINATTGNANITTKTGEINGEVKSASG-NVNITASGNLNVSNITG 821  
Qy 635 --VTSIANGSGSSTNTELKAVLGLTVPNTIPDQGIIMTVKINKDSPAQACAGNOTILCI 692  
Db 822 QNVTVTANSAGTITTE-----GSTI--NATTGDANITTOTGNIINGKVESSGSVT--- 869  
Qy 693 STKNLNTQAMVENGOTLIVGGIYEDNGNTLT 724  
Db 870 -----LIATGOTLAVGNI-----SGDTVT 888

RESULT 40  
US-09-206-942-34  
; Sequence 34, Application US/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; TITLE OF INVENTION: Molecular Weight Proteins  
; FILE REFERENCE: 1038-861 MIS:jb  
; CURRENT APPLICATION NUMBER: US/09/206,942  
; EARLIER FILING DATE: 1998-12-08  
; EARLIER FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 34  
; LENGTH: 1228  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-206-942-34

Query Match 3.7%; Score 141.5; DB 4; Length 1228;  
Best Local Similarity 21.4%; Pred. No. 0.016;  
Matches 135; Conservative 79; Mismatches 201; Indels 217; Gaps 33;  
Qy 239 IIELAALGFAGQPD-----SQQDHIIIVTLKNHT-----LPTTLQSL-- 277  
Db 334 IIESQNFASGSSSLKFKSEGSTHAFTIKNDLIINATGNIISLNOVAGIDSNLKSLIA 393  
Qy 278 -----DVADFKTPVQ-----KVTLKRLN--NDTQLI-----ITAGNW 308  
Db 394 NKNITFEGNITLAADKKPIEIKGNTVKEGANVTLRSANYGNDKSAISIRGNVTNKNL 453  
Qy 309 ELVNKS-----AAPGYFTFQVLPPKQNLSESGGVNAPKFTTGRKISLD-----FQDVE 356  
Db 454 TVTGSAINIEKNLTVEGSAKFLANPNYSFNVSLFDNQGKS---NISIAGGAIFKDI 509

```
QY 357 IRTILOLAKESGNIIVASDS-----VNGKWTLSLKDVDPWQOALDLVQCARLDMRQGN 411
Db 510 NTGSLNITTK-----SDSNHHTIIKGNITNRKD-----LNI TNNGDNTEIQGN 555
QY 412 I-----VNIAPRELLAKDAPLOAEKDIDLALYSONFOLKYK-----451
Db 556 ISOKEGNTLITSSDKVITERITIKA-----GVNGDSDSNEATSANLTIKTKEKLND 609
QY 452 -NVEEPR--SILRLNADTTGNENTLVSGRGSVLIDPATNTLIVTDRSVIEKPKLID- 507
Db 610 LNISGFNKAEITAKDNSNLTIGDN---SDAGN-----TDAKKV--TFSNVKDS 652
QY 508 ELDPAQOQVMIERIVERAADGFSRDLG-----VKFCATGKKKL-----KN 547
Db 653 KISASDHNVTLNKSQVETSGDTEDEGNNNTGLITAKNVTWNNNITSHKTVNTASEN 712
QY 548 DTSFAGWGVNSGFG-----GDDKWGAETKI-NLPITAAANSISLVR-----AIS 590
Db 713 VTTKAGTTINATTGSVEVTAKTGDIKGGIESNSGNVNTITASGDTLNVSNITGQNVTVAAA 772
QY 591 SGALNLELSASELSKTKTLANPRVLTQNRK---EAKIESGYEIPPT-----634
Db 773 SGA-----VTTTGSTINATTGNANITTKTGEINGEVKSASG-NVNITASGNTLNVSNITG 827
QY 635 --VTSIANGGSSTNTELKKAVALGLVTNPNTPDQIIMTVKINKDSPAQOASGNQITLCI 692
Db 828 QNVTVTANSCAITTE-----GSTI--NATTGDANITTTGNGKVESSSGSVT----875
QY 693 STKNLNTQAMVENGTLIVGGIYEEDNGNTLT 724
Db 876 -----LIATGQTLAVGNI-----SGDTVT 894
```

Search completed: December 9, 2003, 10:28:23  
Job time : 30 secs

M catarrhalis MCA1  
Pseudomonas XcpQ s  
Pseudomonas alcali  
Pseudomonas alcali  
Pseudomonas alcali  
E. coli cellular p  
Pseudomonas aerugi  
Chlamydia trachoma  
Bordetella pertuss  
Chlamydia pneumonia  
Chlamydia trachoma  
C. pneumoniae CPN1  
Chlamydia pneumonia  
Chlamydia pneumonia  
Listeria monocytog  
Novel human diagno  
Pathogen specific  
Staphylococcus aur  
Staphylococcus aur  
Staphylococcus aur  
S. aureus antigeni  
Staphylococcus aur  
Novel human diagno  
Novel human diagno  
Haemophilus influe  
Haemophilus influe  
Staphylococcus aur  
Staphylococcus aur  
Staphylococcus aur  
Amino acid sequenc  
Haemophilus influe  
High molecular wei  
Haemophilus high m  
Non-typeable Haemo  
Haemophilus influe

|    |       |      |       |    |          |
|----|-------|------|-------|----|----------|
| 10 | 664.5 | 17.3 | 473   | 23 | AA017573 |
| 11 | 341.5 | 8.9  | 649   | 19 | AAW53827 |
| 12 | 341.5 | 8.9  | 649   | 21 | AAW82594 |
| 13 | 341.5 | 8.9  | 649   | 22 | AAW82256 |
| 14 | 341.5 | 8.9  | 649   | 23 | AAE13660 |
| 15 | 332   | 8.6  | 654   | 22 | AAU34768 |
| 16 | 300   | 7.8  | 776   | 22 | AAU33596 |
| 17 | 278   | 7.2  | 666   | 20 | AAU37501 |
| 18 | 250.5 | 6.5  | 600   | 21 | AAU14113 |
| 19 | 233   | 6.1  | 561   | 20 | AAU35326 |
| 20 | 229.5 | 6.0  | 701   | 20 | AAU37640 |
| 21 | 226   | 5.9  | 698   | 21 | AAU92827 |
| 22 | 212.5 | 5.5  | 754   | 20 | AAU35456 |
| 23 | 212.5 | 5.5  | 754   | 23 | ABB90607 |
| 24 | 183   | 4.8  | 1711  | 23 | ABB47325 |
| 25 | 182   | 4.7  | 671   | 22 | ABG18141 |
| 26 | 167.5 | 4.4  | 10498 | 24 | ABJ19119 |
| 27 | 167   | 4.3  | 1029  | 22 | AAU34389 |
| 28 | 166.5 | 4.3  | 1048  | 22 | AAU37490 |
| 29 | 166.5 | 4.3  | 2086  | 22 | AAU34143 |
| 30 | 166.5 | 4.3  | 2659  | 23 | AAU75489 |
| 31 | 166.5 | 4.3  | 5795  | 22 | AAU37017 |
| 32 | 165   | 4.3  | 5464  | 22 | ABG11810 |
| 33 | 185   | 4.3  | 7718  | 22 | ABG11811 |
| 34 | 160.5 | 4.2  | 1073  | 21 | AAU01837 |
| 35 | 160.5 | 4.2  | 1079  | 21 | AAU01836 |
| 36 | 159.5 | 4.1  | 6281  | 22 | AAU37403 |
| 37 | 159   | 4.1  | 837   | 22 | AAU34387 |
| 38 | 159   | 4.1  | 875   | 22 | AAU37487 |
| 39 | 158   | 4.1  | 2434  | 22 | AAU34339 |
| 40 | 156.5 | 4.1  | 1978  | 20 | AAU27230 |
| 41 | 155.5 | 4.0  | 1095  | 21 | AAU01847 |
| 42 | 155.5 | 4.0  | 1536  | 14 | AAU41723 |
| 43 | 155.5 | 4.0  | 1536  | 14 | AAU63505 |
| 44 | 155.5 | 4.0  | 1536  | 18 | AAU30293 |
| 45 | 155.5 | 4.0  | 1536  | 21 | AAU01846 |

ALIGNMENTS

RESULT 1  
AAU53895  
ID AAY53895 standard; Protein; 769 AA.  
AC AAY53895;  
XX  
XX 13-MAR-2000 (first entry)  
XX A Neisseria meningitidis antigenic protein designated PASB030.  
XX Antigenic polypeptide: PASB030; serotype B strain ATCC 13090;  
XX vaccine; infection; bacteremia; meningitis.  
XX Neisseria meningitidis.  
OS  
PN WO9961620-A2.  
PN  
XX  
PD 02-DEC-1999.  
XX  
PP 26-MAY-1999; 99WO-EP03603.  
XX  
PR 26-MAY-1998; 98GB-0011260.  
XX  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
XX Ruelle J;  
XX WPI; 2000-072624/06.  
XX N-PSDB; AAZ36857.  
XX New isolated Neisseria meningitidis polypeptides and polynucleotides,  
PT used to develop products for the diagnosis, prevention and treatment of

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 10:19:37 ; Search time 48 Seconds  
(without alignments)

2542.932 Million cell updates/sec

Title: US-09-701-271A-2

Perfect score: 3848

Sequence: 1 MNTKLTIIISGLFVAFAFO.....ELLIFITPRIMGTAGNSLRY 769

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 3848  | 100.0       | 769    | 21 | AAY53895    |
| 2          | 3723  | 96.8        | 769    | 21 | AAY53896    |
| 3          | 3707  | 96.3        | 769    | 21 | AAY53897    |
| 4          | 3482  | 90.5        | 723    | 24 | ABP77693    |
| 5          | 2988  | 77.7        | 720    | 23 | ABG91062    |
| 6          | 664.5 | 17.3        | 473    | 21 | AAU44394    |
| 7          | 664.5 | 17.3        | 473    | 21 | AAU44395    |
| 8          | 664.5 | 17.3        | 473    | 21 | AAU44396    |
| 9          | 664.5 | 17.3        | 473    | 21 | AAU44397    |

infections -

XX Claim 5; Page 86-88; 97pp; English.

XX The present sequence represents a *Neisseria meningitidis* antigenic polypeptide, designated BASB030. It was identified from *N. meningitidis* serotype B strain ATCC 13090. The nucleotide sequence was first identified in the Incyte Pathoseq database of unfinished genomic DNA sequences of this strain. The polypeptides or polynucleotides can be used in vaccine compositions for preventing *N. meningitidis* infections, e.g. bacteremia and meningitis. The antibodies against the protein can be used for treating *N. meningitidis* disease. The products can also be used for diagnosis of disease, staging of disease or response of an infectious organism to drugs, as well as for drug screening.

SQ Sequence 769 AA;

Query Match 100.0%; Score 3848; DB 21; Length 769;  
Best Local Similarity 100.0%; Pred. No. 9.2e-278; Indels 0; Gaps 0;  
Matches 769; Conservative 0; Mismatches 0;

QY 1 MNTKLTIIISGLFVATAAFQTASAGNITDIKVSSLPNKQIKVSVFDEIYNPTGFTVSS 60  
Db 1 MNTKLTIIISGLFVATAAFQTASAGNITDIKVSSLPNKQIKVSVFDEIYNPTGFTVSS 60

QY 61 PARIALDFEQTGISMDQOVLEYADPLLSKISAQNSRRARLVNLNKPQYNTVEVRGNKV 120  
Db 61 PARIALDFEQTGISMDQOVLEYADPLLSKISAQNSRRARLVNLNKPQYNTVEVRGNKV 120

QY 121 WIFINESDDTVSAPARPAVKAAPAAKQOQCRVTYQVRSIRIOTLYPGKTTAAAPFTES 180  
Db 121 WIFINESDDTVSAPARPAVKAAPAAKQOQCRVTYQVRSIRIOTLYPGKTTAAAPFTES 180

QY 181 VVVSAPSPSPAKQAAASAKOQTAAKQOQCRVTYQVRSIRIOTLYPGKTTAAAPFTES 240  
Db 181 VVVSAPSPSPAKQAAASAKOQTAAKQOQCRVTYQVRSIRIOTLYPGKTTAAAPFTES 240

QY 241 ELAALGFAGQPDISOQHDHIIIVTLKNHTLPTTLQSLDVADEKTPVQKVLKRLNNDTQL 300  
Db 241 ELAALGFAGQPDISOQHDHIIIVTLKNHTLPTTLQSLDVADEKTPVQKVLKRLNNDTQL 300

QY 301 IITAGNWLKNSAAGPYFTQVLPKQNLSEGGVNNAPKFTTGRKISLDFQDVEIRTI 360  
Db 301 IITAGNWLKNSAAGPYFTQVLPKQNLSEGGVNNAPKFTTGRKISLDFQDVEIRTI 360

QY 361 LQILAKESGMNIVASDSVNGKMTLSLKDVPWQALDVLVQARNLDMRQGNIVNIAPRDE 420  
Db 361 LQILAKESGMNIVASDSVNGKMTLSLKDVPWQALDVLVQARNLDMRQGNIVNIAPRDE 420

QY 421 LLAOKAFLOAEKDIALDGLYSQNFQLYKQVVEFRSILRLDNADTTGNRTLVSGRGS 480  
Db 421 LLAOKAFLOAEKDIALDGLYSQNFQLYKQVVEFRSILRLDNADTTGNRTLVSGRGS 480

QY 481 VLIDPATNTLIVTDRSVIEKFRKLDIDLVPAQQVMIEARIVEAADGFSRDLGVKFGAT 540  
Db 481 VLIDPATNTLIVTDRSVIEKFRKLDIDLVPAQQVMIEARIVEAADGFSRDLGVKFGAT 540

QY 541 GKXKLKNDTSAPFGWVNSGFGDDKWAETKINLPITAAANSISLVRAISSGALNLELSA 600  
Db 541 GKXKLKNDTSAPFGWVNSGFGDDKWAETKINLPITAAANSISLVRAISSGALNLELSA 600

QY 601 SESLSKTKTLANPRVLTONKEAKIESGVEIPFTVTSIANGSGSTTELKXAVGLTVTP 660  
Db 601 SESLSKTKTLANPRVLTONKEAKIESGVEIPFTVTSIANGSGSTTELKXAVGLTVTP 660

QY 661 NITPDGQIIMTVKINKDSPACQASGNTILCISTKNLNTQAMVENGGLTIVGIVEEDNG 720  
Db 661 NITPDGQIIMTVKINKDSPACQASGNTILCISTKNLNTQAMVENGGLTIVGIVEEDNG 720

QY 721 NTLTKVPLLDGIPIVGNLFKTRGKKTDRRELLFIITPRIMGTAGNSLRY 769  
Db 721 NTLTKVPLLDGIPIVGNLFKTRGKKTDRRELLFIITPRIMGTAGNSLRY 769

RESULT 2

AAV53896

ID AAV53896 standard; Protein; 769 AA.

XX AAY53896;

AC AAY53896;

XX 13-MAR-2000 (first entry)

DE A *Neisseria meningitidis* antigenic protein designated BASB030.

XX Antigenic polypeptide; BASB030; serotype B strain ATCC 13090;

KW vaccine; infection; bacteremia; meningitis.

XX *Neisseria meningitidis*.

PN WO9961620-A2.

XX 02-DEC-1999.

XX 26-MAY-1999; 99WO-EP03603.

XX 26-MAY-1998; 98GB-0011260.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PA Ruelle J;

PI WPI; 2000-072624/06.

DR N-PSDB; AAZ36856.

XX New isolated *Neisseria meningitidis* polypeptides and polynucleotides, of used to develop products for the diagnosis, prevention and treatment of infections.

PS Claim 3; Page 90-92; 97pp; English.

XX The present sequence represents a *Neisseria meningitidis* antigenic polypeptide, designated BASB030. It was identified from *N. meningitidis* serotype B strain ATCC 13090. The nucleotide sequence was first identified in the Incyte Pathoseq database of unfinished genomic DNA sequences of this strain. The polypeptides or polynucleotides can be used in vaccine compositions for preventing *N. meningitidis* infections, e.g. bacteremia and meningitis. The antibodies against the protein can be used for treating *N. meningitidis* disease. The products can also be used for diagnosis of disease, staging of disease or response of an infectious organism to drugs, as well as for drug screening.

SQ Sequence 769 AA;

Query Match 96.8%; Score 3723; DB 21; Length 769;  
Best Local Similarity 97.4%; Pred. No. 1.9e-268; Indels 0; Gaps 0;  
Matches 749; Conservative 2; Mismatches 18;

QY 1 MNTKLTIIISGLFVATAAFQTASAGNITDIKVSSLPNKQIKVSVFDEIYNPTGFTVSS 60  
Db 1 MNTKLTIIISGLFVATAAFQTASAGNITDIKVSSLPNKQIKVSVFDEIYNPTGFTVSS 60

QY 61 PARIALDFEQTGISMDQOVLEYADPLLSKISAQNSRRARLVNLNKPQYNTVEVRGNKV 120  
Db 61 PARIALDFEQTGISMDQOVLEYADPLLSKISAQNSRRARLVNLNKPQYNTVEVRGNKV 120

QY 121 WIFINESDDTVSAPARPAVKAAPAAKQOQCRVTYQVRSIRIOTLYPGKTTAAAPFTES 180  
Db 121 WIFINESDDTVSAPARPAVKAAPAAKQOQCRVTYQVRSIRIOTLYPGKTTAAAPFTES 180

QY 181 VVVSAPSPSPAKQAAASAKOQTAAKQOQCRVTYQVRSIRIOTLYPGKTTAAAPFTES 240  
Db 181 VVVSAPSPSPAKQAAASAKOQTAAKQOQCRVTYQVRSIRIOTLYPGKTTAAAPFTES 240

QY 241 ELAALGFAGQPDISOQHDHIIIVTLKNHTLPTTLQSLDVADEKTPVQKVLKRLNNDTQL 300  
Db 241 ELAALGFAGQPDISOQHDHIIIVTLKNHTLPTTLQSLDVADEKTPVQKVLKRLNNDTQL 300

5/26/75

CC identified in the Incyte Pathoseq database of unfinished genomic DNA  
sequences of this strain. The polypeptides or polynucleotides can be  
used in vaccine compositions for preventing N. meningitidis infections,  
e.g. bacteremia and meningitis. The antibodies against the protein can  
be used for treating N. meningitidis disease. The products can also  
be used for diagnosis of disease, staging of disease or response of an  
infectious organism to drugs, as well as for drug screening.

XX Sequence 769 AA;

QY 301 IITAGNWLKNSAAGPYFTFQVLPKQKQNSGGVNNAPKFTGRKISLDFQDVEIRTI 360  
DB 301 IITAGNWLKNSAAGPYFTFQVLPKQKQNSGGVNNAPKFTGRKISLDFQDVEIRTI 360  
QY 361 LQILAKESGMNIVASDSVNGKMTLSLKDVPDQALDLVMOARLDMRQGNINVIAPRDE 420  
DB 361 LQILAKESGMNIVASDSVNGKMTLSLKDVPDQALDLVMOARLDMRQGNINVIAPRDE 420  
QY 421 LLAKDKAFLQAEKDIALDGLALYSONFOLKYKNVEEFSRLRLDNADTTGNRNTLVSGRGS 480  
DB 421 LLAKDKAFLQAEKDIALDGLALYSONFOLKYKNVEEFSRLRLDNADTTGNRNTLVSGRGS 480  
QY 481 VLIDPATNTLIVTDRSVIEKFKRLIDELDVPQQVMI EARIVEAADGFSRDLGVKFGAT 540  
DB 481 VLIDPATNTLIVTDRSVIEKFKRLIDELDVPQQVMI EARIVEAADGFSRDLGVKFGAT 540  
QY 541 GKXKLKNDTSAPFGWVNSGFGGDDKWAETKINLPITAAANSISLVRAISSGALNLELSA 600  
DB 541 GKXKLKNDTSAPFGWVNSGFGGDDKWAETKINLPITAAANSISLVRAISSGALNLELSA 600  
QY 601 SESLSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANGGSSNTTELKKAIVGLTVP 660  
DB 601 SESLSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANGGSSNTTELKKAIVGLTVP 660  
QY 661 NITPDGQIIMTVKINKDSPACASGNQTLICISTKNLNTQAMVENGTLIVGGIYEDNG 720  
DB 661 NITPDGQIIMTVKINKDSPACASGNQTLICISTKNLNTQAMVENGTLIVGGIYEDNG 720  
QY 721 NTLTKVPLLGDIPIVGNLFKTRGKKTDRRELLIFITPRIMGTAGNSLRY 769  
DB 721 NTLTKVPLLGDIPIVGNLFKTRGKKTDRRELLIFITPRIMGTAGNSLRY 769

RESULT 3  
AAV53897  
ID AAV53897 standard; Protein; 769 AA.  
XX AAV53897;  
XX 13-MAR-2000 (first entry)  
XX A Neisseria meningitidis antigenic protein designated BASB030.  
XX Antigenic polypeptide; BASB030; serotype B strain H44/76;  
XX vaccine; infection; bacteremia; meningitis.  
XX Neisseria meningitidis.  
XX WO9961C20-A2.  
XX 02-DEC-1999.  
XX 26-MAY-1999; 99WO-EP03603.  
XX 26-MAY-1998; 98GB-0011260.  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX Ruelle J;  
XX WPI: 2000-072624/06.  
XX N-PSDB; AAZ36859.  
XX New isolated Neisseria meningitidis polypeptides and polynucleotides,  
XX used to develop products for the diagnosis, prevention and treatment of  
XX infections -  
XX Claim 3; Page 93-96; 97pp; English.  
XX The present sequence represents a Neisseria meningitidis antigenic  
XX polypeptide, designated BASB030. It was identified from N. meningitidis  
XX serotype B strain H44/76. The nucleotide sequence was first

CC sequences of this strain. The polypeptides or polynucleotides can be  
used in vaccine compositions for preventing N. meningitidis infections,  
e.g. bacteremia and meningitis. The antibodies against the protein can  
be used for treating N. meningitidis disease. The products can also  
be used for diagnosis of disease, staging of disease or response of an  
infectious organism to drugs, as well as for drug screening.

XX Sequence 769 AA;

QY 1 MNTKLTIIISGLFVATAAFAQTASAGNIITDIKVSSLPNKQIKVKSFDKEIVNPTGFTVSS 60  
DB 1 MNTKLTIIISGLFVATAAFAQTASAGNIITDIKVSSLPNKQIKVKSFDKEIVNPTGFTVSS 60  
QY 61 PARIALDEPQTGISMDQVQVLEYADPPLLKISAAQNSRRARLVNLNPKQYNTVRGNKV 120  
DB 61 PARIALDEPQTGISMDQVQVLEYADPPLLKISAAQNSRRARLVNLNPKQYNTVRGNKV 120  
QY 121 WIFINESDDTVSAPARPAPKAPPAKQOQCRITVYVRSIRIQTLYPGKTTAAAPFTES 180  
DB 121 WIFINESDDTVSAPARPAPKAPPAKQOQCRITVYVRSIRIQTLYPGKTTAAAPFTES 180  
QY 181 VVSVAAPPSPAKQQAASAKQQTAAAPAKQQAAPAKQQTNIIDFRKDGKNAGII 240  
DB 181 VVSVAAPPSPAKQQAASAKQQAAPAKQQAAPAKQQTNIIDFRKDGKNAGII 240  
QY 241 ELAALGFAGQDPSIQOHDHIIIVTLKNTLPTTLQSLDVADEKTPVQVTLKRLNNDTQL 300  
DB 241 ELAALGFAGQDPSIQOHDHIIIVTLKNTLPTTLQSLDVADEKTPVQVTLKRLNNDTQL 300  
QY 301 IITAGNWLKNSAAGPYFTFQVLPKQKQNSGGVNNAPKFTGRKISLDFQDVEIRTI 360  
DB 301 IITAGNWLKNSAAGPYFTFQVLPKQKQNSGGVNNAPKFTGRKISLDFQDVEIRTI 360  
QY 361 LQILAKESGMNIVASDSVNGKMTLSLKDVPDQALDLVMOARLDMRQGNINVIAPRDE 420  
DB 361 LQILAKESGMNIVASDSVNGKMTLSLKDVPDQALDLVMOARLDMRQGNINVIAPRDE 420  
QY 421 LLAKDKAFLQAEKDIALDGLALYSONFOLKYKNVEEFSRLRLDNADTTGNRNTLVSGRGS 480  
DB 421 LLAKDKAFLQAEKDIALDGLALYSONFOLKYKNVEEFSRLRLDNADTTGNRNTLVSGRGS 480  
QY 481 VLIDPATNTLIVTDRSVIEKFKRLIDELDVPQQVMI EARIVEAADGFSRDLGVKFGAT 540  
DB 481 VLIDPATNTLIVTDRSVIEKFKRLIDELDVPQQVMI EARIVEAADGFSRDLGVKFGAT 540  
QY 541 GKXKLKNDTSAPFGWVNSGFGGDDKWAETKINLPITAAANSISLVRAISSGALNLELSA 600  
DB 541 GKXKLKNDTSAPFGWVNSGFGGDDKWAETKINLPITAAANSISLVRAISSGALNLELSA 600  
QY 601 SESLSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANGGSSNTTELKKAIVGLTVP 660  
DB 601 SESLSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANGGSSNTTELKKAIVGLTVP 660  
QY 661 NITPDGQIIMTVKINKDSPACASGNQTLICISTKNLNTQAMVENGTLIVGGIYEDNG 720  
DB 661 NITPDGQIIMTVKINKDSPACASGNQTLICISTKNLNTQAMVENGTLIVGGIYEDNG 720  
QY 721 NTLTKVPLLGDIPIVGNLFKTRGKKTDRRELLIFITPRIMGTAGNSLRY 769  
DB 721 NTLTKVPLLGDIPIVGNLFKTRGKKTDRRELLIFITPRIMGTAGNSLRY 769

RESULT 4  
ABP77693  
ID ABP77693 standard; Protein; 723 AA.  
XX ABP77693;  
XX

07-MAR-2003 (first entry)  
N. gonorrhoeae amino acid sequence SEQ ID 1916.  
Antibacterial; infection; vaccine; gene therapy.  
Neisseria gonorrhoeae.  
WO200279243-A2.  
10-OCT-2002.  
12-FEB-2002; 2002WO-IB02069.  
12-FEB-2001; 2001GB-0003424.  
(CHIR-) CHIRON SPA.  
Fontana MR, Pizza M, Massignani V, Monaci E;  
WPI; 2003-058415/05.  
DR N-PSDB; ABZ38663.  
New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
medicament for treating or preventing N. gonorrhoeae infection -  
Disclosure; Page 326; 815pp; English.  
The present invention relates to proteins from Neisseria gonorrhoeae.  
Also disclosed are the nucleic acid molecules encoding the proteins and  
antibodies that specifically bind to the proteins. The composition  
comprising the protein, nucleic acid or antibody is useful for the  
manufacture of a medicament for treating or preventing N. gonorrhoeae  
infection, this may be in the form of a vaccine or gene therapy.  
Sequences given in records ABP7636-ABP81046 represent nucleic acid  
molecules of the invention.  
Sequence 723 AA;  
Query Match 90.5%; Score 3482; DB 24; Length 723;  
Best Local Similarity 91.4%; Pred. No. 1.6e-250;  
Matches 703; Conservative 11; Mismatches 9; Indels 46; Gaps 2;  
QY 1 MNTKLTIIISGLFVATAAFQTASAGNITDIKVSSLPNKKIVKVSFDEIVNPTGFVTSS 60  
DB 1 MNTKLTIIISGLFVATAAFQTASAGNITDIKVSSLPNKKIVKVSFDEIVNPTGFVTSS 60  
QY 61 PARIADFEQTGISMDOQVLEADPLLSKISAAQSSRRARLVNLNPKQYNTVEVRGNKV 120  
DB 61 PARIADFEQTGISMDOQVLEADPLLSKISAAQSSRRARLVNLNPKQYNTVEVRGNKV 120  
QY 121 WIFINESDDTVSAPAPVAKAAPAAKQOQTAAAPAKQAAAPAKQO-----AAAPFTES 180  
DB 121 WIFINESDDTVSAPAPVAKAAPAAKQO-----AAAPFTES 158  
QY 181 VVSVSAPFSPAKQAAASAKQOQTAAAPAKQAAAPAKQO-----AKQINIDFRKDGKNAGII 240  
DB 159 VVSVSAPFSPAKQAAAS-----AKQINIDFRKDGKNAGII 194  
QY 241 ELAALGFAGQPDISQOHDHIIIVLKNHTLPTLQSRSLDVADFKTPVQKVTLELNNDTOL 300  
DB 195 ELAALGFAGQPDISQOHDHIIIVLKNHTLPTLQSRSLDVADFKTPVQKVTLELNNDTOL 254  
QY 301 IITAGNWLWLNKSAAPGYFTFOVLPKKQNLSEGGVNNAPKFTTGRKISLDFQDVEIRTI 360  
DB 255 IITAGNWLWLNKSAAPGYFTFOVLPKKQNLSEGGVNNAPKFTTGRKISLDFQDVEIRTI 314  
QY 361 LQILAKESGNMIVASDSVNGKMTLSLKDVPDQALDVMQARNLDMRQGNIVNIAPRDE 420  
DB 315 LQILAKESGNMIVASDSVNGKMTLSLKDVPDQALDVMQARNLDMRQGNIVNIAPRDE 374  
QY 421 LIAKDAFLQAEKDIALGLALYQNQLKYKNVEFRSILRLDNADTTGNRNTLVSGRGS 480

DB 375 LIAKDAFLQAEKDIALGLALYQNQLKYKNVEFRSILRLDNADTTGNRNTLVSGRGS 434  
QY 481 VLIDPATNTLIVTDRSVIEKFRKIDELDVPAQQVMIEARIVEAADGFSRDLGVKFGAT 540  
DB 435 VLIDPATNTLIVTDRSVIEKFRKIDELDVPAQQVMIEARIVEAADGFSRDLGVKFGAT 494  
QY 541 GKKKLNDTSAFQGWVNSGFGGDDKGAETKINLPITAAANSISLVAISSGALNLELSA 600  
DB 495 GRKKLXNETSAFQGWVNSGFGGDKWEAQTKINLPVAAAANSISLVAISSGALNLELSA 554  
QY 601 SESLSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANGSGSSTNTEKKAVLGLTVTP 660  
DB 555 SESLSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANGSGSSTNTEKKAVLGLTVTP 614  
QY 661 NITPDGQIIMTVKINKDSPAQCASGNQTLICISTKNLNTQAMVNGGTLIVGGIYEEDNG 720  
DB 615 NITPDGQIIMTVKINKDSPAQCASGNNTILCISTKNLNTQAMVNGGTLIVGGIYEENNG 674  
QY 721 NTLTQVPLLDGDIPIVGNLFRKTRKKTDRRELLIFITPRIMGTAGNSLRY 769  
DB 675 NTLTQVPLLDGDIPIVGNLFRKTRKKTDRRELLIFITPRIIDTAGNSLRY 723

RESULT 5  
ABG91062  
ID ABG91062 standard; Protein; 720 AA.  
XX AC ABG91062;  
XX DT 29-NOV-2002 (first entry)  
XX DE Neisseria gonorrhoeae pilus biogenesis protein #3.  
XX KW Gram-negative bacterial bleb; PorB; outer membrane protein;  
XX KW Chlamydia trachomatis infection; Chlamydia pneumoniae infection;  
XX KW Protective antigen; antibacterial; vaccine.  
XX OS Neisseria gonorrhoeae.  
XX PN WO200262380-A2.  
XX PD 15-AUG-2002.  
XX PF 08-FEB-2002; 2002WO-EP01356.  
XX PR 08-FEB-2001; 2001GB-0003169.  
XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.  
XX PI Berthet FJ, Lobet Y, Poolman J, Verlant VGCL;  
XX DR WPI; 2002-657510/70.  
XX DR N-PSDB; ABS67380.  
XX PT Novel gram-negative bacterial bleb presenting on its surface PorB outer  
PT membrane protein from Chlamydia trachomatis or protective antigen from  
PT Chlamydia pneumoniae, useful for preventing Chlamydia infection -  
XX PS Disclosure; Page 52; 75pp; English.  
XX CC The present invention relates to a new gram-negative bacterial bleb  
XX CC presenting on its surface the PorB outer membrane protein from Chlamydia  
XX CC trachomatis, or a protective antigen from C. pneumoniae. The invention  
XX CC is useful for preventing C. trachomatis or C. pneumoniae infection in a  
XX CC host. The present amino acid sequence represents a Neisseria gonorrhoeae  
XX CC protein as described in the invention.  
XX SQ Sequence 720 AA;  
Query Match 77.7%; Score 2988; DB 23; Length 720;  
Best Local Similarity 81.7%; Pred. No. 9.7e-214;  
Matches 619; Conservative 21; Mismatches 78; Indels 40; Gaps 4;

QY 1 MNTKLTIIISGLFVATAAFOTASAGNITDIKVSSLPNKQIVKVSFDFKEIVNPTGFTVSS 60  
DB 1 MNTKLTIIISGLFVATAAFOTASAGNITDIKVSSLPNKQIVKVSFDFKEIVNPTGFTVSS 60  
QY 61 PARIALDFEQTGISMDQVLEVADPLLSKISAAQNSSRARLVNLNPKQYNTVEVGNKV 120  
DB 61 PARIALDFEQTGISMDQVLEVADPLLSKISAAQNSSRARLVNLNPKQYNTVEVGNKV 120  
QY 121 WIFINESDDTVSAPAPVAKAAPAKQOQCKTVVQVRSIRIQTLYPGKTTAAAPFTES 180  
DB 121 WIFINESDDTVSAPAPVAKAAPAKQOQCKTVVQVRSIRIQTLYPGKTTAAAPFTES 158  
QY 181 VVSVSAPFSPAKQOAAASAKOQTAAPAKOQAAAPAKQTNIDFRDKGNAGII 240  
DB 159 VVSVSAPFSPAKQOAAASAKOQTAAPAKOQAAAPAKQTNIDFRDKGNAGII 202  
QY 241 ELAALFAGOPDISOQHDHIIIVTLKNHTLPTTLORSLDVADPKTPVKQVTKLRNLNDTOL 300  
DB 203 ELAALFAGOPDISOQHDHIIIVTLKNHTLPTALORSLDVADPKTPVKQVTKLRNLNDTOL 262  
QY 301 IITTAGNWLNVKSAAPGYFTFOVLPPKQNLSEGGVNNAPKFTTGRKISLDFQDVEIRTI 360  
DB 263 IITTAGNWLNVKSAAPGYFTFOVLPPKQNLSEGGVNNAPKFTTGRKISLDFQDVEIRTI 322  
QY 361 LQILAKESGMNIVASDSVNGKMTLSLKDVDPDOALDVMQARNLDMROQGNIVNAPRDE 420  
DB 323 LQILAKESGMNIVASDSVNGKMTLSLKDVDPDOALDVMQARNLDMROQGNIVNAPR-R 381  
QY 421 LLAQKAFLOAEKDIAADLGALYSQNFQYKNVEEFSRLDNDADTTGNRNTLVSGRS 480  
DB 382 AACQROSLTSGRCHRSRAVFPKLPKIEOKCGRIPQHPALDNDADTTGNRNTLVSGRS 441  
QY 481 VLIDPATNLTIVDTRSVIEKFKRLIDELVPAQOQVMEARIVEAADGFSRDLGVKFGAT 540  
DB 442 VLIDPATNLTIVDTRSVIEKFKRLIDELVPAQOQVMEARIVEAADGFSRDLGVKFGAT 501  
QY 541 GKXKLKNDTSFAGWVNSGFGGDKWGAETKINLPITAAANSISLVRAISSGALNLELSA 600  
DB 502 GRKXKLKNDTSFAGWVNSGFGGDKWGAETKINLPITAAANSISLVRAISSGALNLELSA 560  
QY 601 SESLSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANGSGSSTNTELLKAVLGLTVTP 660  
DB 561 SESLSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANGSGSSTNTELLKAVLGLTVTP 620  
QY 661 NITPDGQIIMTVKINKDSPAQCSAGNQTILCISTKNLNTQAMVENGGLTVGGIYEEDNG 720  
DB 621 NITPDGQIIMTVKINKDSPAQCSAGNQTILCISTKNLNTQAMVENGGLTVGGIYEEDNG 680  
QY 721 NTLTKVPLLDGIPVIGNLFTKRGKTKDRRELLIFITPR 758  
DB 681 NTLTKVPLLDGIPVIGNLFTKRGKTKDRRELLIFITPR 718  
RESULT 6  
ID AAY44394 standard; Protein; 473 AA.  
XX  
AC AAY44394;  
XX  
XX 22-MAR-2000 (first entry)  
XX  
DE Moraxella catarrhalis BAS031 protein-1, from strain Mc2931 (ATCC 43617).  
XX  
KW BAS031; strain Mc2931; ATCC 43617; PilQ fimbrial assembly protein;  
KW diagnosis; treatment; otitis media; sinusitis; pneumonia; screening;  
KW homology; nosocomial infection; antibody; ortholog; hybridisation probe.  
XX  
OS Moraxella catarrhalis.  
XX  
PN W09964448-A2.  
XX  
PD 16-DEC-1999.  
XX

PF 31-MAY-1999; 99WO-EF03823.  
XX  
PR 05-JUN-1998; 98GB-0012163.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
PI Ruelle J, Tommassen JPM, Vinals-Bassols C;  
XX  
DR WPI; 2000-116523/10.  
XX N-PSDB; AAZ29556.  
PT Novel polypeptides used as vaccines for treating Maraxella catarrhalis  
XX infections like otitis media and pneumonia  
PS Claim 5; Page 97-98; 121pp; English.  
XX  
CC The present sequence is the BAS031 protein sequence-1, from strain  
CC Mc2931(ATCC 43617), derived from Moraxella catarrhalis. This polypeptide  
CC sequence has homology to Pseudomonas aeruginosa, PilQ fimbrial assembly  
CC protein. This sequence can be used for prevention and treatment of  
CC M.catarrhalis infections, like otitis media, pneumonia, sinusitis and  
CC nosocomial infections. The antibodies and polynucleotide sequence can be  
CC used for diagnosing infections, staging of disease and for determining  
CC the response of an infectious organism to drugs. The DNA sequence is  
CC also used as a hybridisation probe in screening process for identifying  
CC homologue and orthologs from other species.  
XX  
SQ Sequence 473 AA;  
Query Match 17.3%; Score 664.5; DB 21; Length 473;  
Best Local Similarity 34.4%; Pred. No. 6.4e-41;  
Matches 158; Conservative 87; Mismatches 153; Indels 61; Gaps 12;  
QY 339 APKFTGKISLDFQDVEIRTIQILAKESGMNIVASDSVNGKMTLSLKDVDPDOALDVLV 398  
DB 38 AEQHYTGKPISEFADIPRAVFDILAGFTGINIITDSDVTGSMITRLNLPWQAPDVI 97  
QY 399 MOARNLDMROQGNIVNAPRDE-----BLAKKAFLOAEKDIAADLGALYSQNFOLK 449  
DB 98 LQTNLSVLKHGWNWLLSSKSIQSNQPTITEVIRNLVALAD--DVATL--IMGEKTQRG 152  
QY 450 YKNVEEFSRL-----RLDNDADTTGNRNTLVSGRSVLIDPATNLTIVDTR 497  
DB 153 NVNRTNHRDVIYPEAPLSRAYNTQTDLEITAVRGTLSSERGTVDKRTNLTIIQDVPA 212  
QY 498 VIEKFKRLIDELVPAQOQVMEARIVEAADGFSRDLGVKFGATGKKLKNDSAPG---- 553  
DB 213 SVANKALIERIDIPVEQVMEARIVSANENFORKLVGSFGAHG-----QNGKTVYGSQG 268  
QY 554 -----WGVNSGFGGDDKWGAETKINLPITAAANSISLVRAISSGALN-----LELSA 601  
DB 269 SLWTWVQEGVAAGGHQ--NSHLNVLDLGVDNAMGRIAF-----GLLNLPDIILDELSAM 320  
QY 602 ESLSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANGSGSSTNTELLKAVLGLTVTPN 661  
DB 321 QAENGEVISTPKVLTAQKQARISSSGLQIPYQETT--HSGAST-TKFEASLILEATPN 377  
QY 662 ITPDQGIIMTVKINKDSPAQCSAGNQTILCISTKNLNTQAMVENGGLTVGGIYEEDNGN 721  
DB 378 ITPDGKIGLKLNIKNGNVPVTLGH----IAQEDAETNVIIDGQTVVGLGIVRTSQNE 433  
QY 722 TLTKVPLLDGIPVIGNLFTKRGKTKDRRELLIFITPRIM 760  
DB 434 GANKVPRGLDMVPLGRFLPHDSKAHDKSELLIFITPKLV 472  
RESULT 7  
ID AAY44395 standard; Protein; 473 AA.  
XX  
AC AAY44395;  
XX  
DT 22-MAR-2000 (first entry)

```

XX DE Moraxella catarrhalis BASB031 protein-2, from strain Mc2931 (ATCC 43617).
XX KW BASB031; strain Mc2931; ATCC 43617; PilQ fimbrial assembly protein;
XX KW diagnosis; treatment; otitis media; sinusitis; pneumonia; screening;
XX KW homology; nosocomial infection; antibody; ortholog; hybridisation probe.
XX OS Moraxella catarrhalis.
XX FN WO9964448-A2.
XX PD 16-DEC-1999.
XX PF 31-MAY-1999; 99WO-EP03823.
XX PR 05-JUN-1998; 98GB-0012163.
XX PA (SMIK ) SMITHLINE BEECHAM BIOLOGICALS.
XX PI Ruelle J, Tommassen JPM, Vinals-Bassols C;
XX DR WPI; 2000-116523/10.
XX DR N-PSDB; AAZ29557.
XX KW Novel polypeptides used as vaccines for treating Maraxella catarrhalis
XX PT infections like otitis media and pneumonia
XX PS Claim 3; Page 99-101; 121pp; English.
XX CC The present sequence is the BASB031 protein sequence-2, from strain
XX CC Mc2931 (ATCC 43617), derived from Moraxella catarrhalis. This polypeptide
XX CC sequence has homology to Pseudomonas aeruginosa, PilQ fimbrial assembly
XX CC protein. This sequence can be used for prevention and treatment of
XX CC M.catarrhalis infections, like otitis media, pneumonia, sinusitis and
XX CC nosocomial infections. The antibodies and polynucleotide sequence can be
XX CC used for diagnosing infections, staging of disease and for determining
XX CC the response of an infectious organism to drugs. The DNA sequence is
XX CC also used as a hybridisation probe in screening process for identifying
XX CC homologue and orthologues from other species.
XX SQ Sequence 473 AA;
    Query Match 17.3%; Score 664.5; DB 21; Length 473;
    Best Local Similarity 34.4%; Pred. NO. 6.4e-41;
    Matches 158; Conservative 87; Mismatches 153; Indels 61; Gaps 12;
QY 339 APKFTTGRKISLDFQDVETIRILOILAKESGMNIVASDSVNGKWTLSLKDVDPDQALDLV 398
DB 38 AEQHYTGKPISEFADIPVRAVFDILAGFTGINITDSDSVTGMTIRLNPWDQAFDVI 97
QY 399 MQARNLDMRQOQGNIVNIAPRD-----ELLAKDRAFLQAEKDIAIDLALYSQNFQLK 449
DB 98 LQTNLSVLKHGVMWLISSKSIQSNQPTITEYIRLNYALAD---DVATL--IMGEKTQRG 152
QY 450 YKNVEFRSIL-----RLDNADTTGNRTLVSGRGSVLIDPATNLTIVTDTRS 497
DB 153 NVNRTNHRDVIYPEAPLSRAYNTQTDELITTAVRGTLTSSRGTVTDKRTNLTIIQDVP 212
QY 498 VIEKFRKILDELDPVPAQOVMEARIVEAADGFSRDLGVKFGATGKK-----K 544

```

```

QY 722 TLTKVPLLGDPVIGNLFKTRGKKTDRRELLIFITPRIM 760
DB 434 GANKVPLRGDMFVJGRLEFRHDSKAHDKSELLIFITPKLV 472
RESULT 8
AAY44396
ID AAY44396 standard; Protein; 473 AA.
XX AC AAY44396;
XX DT 22-MAR-2000 (first entry)
XX DE Moraxella catarrhalis BASB031 protein sequence from strain Mc2911.
XX KW BASB031; strain Mc2911; PilQ fimbrial assembly protein; homology;
XX KW diagnosis; treatment; otitis media; pneumonia; sinusitis; antibody;
XX KW nosocomial infection; screening; hybridisation probe; ortholog.
XX OS Moraxella catarrhalis.
XX FN WO9964448-A2.
XX PD 16-DEC-1999.
XX PF 31-MAY-1999; 99WO-EP03823.
XX PR 05-JUN-1998; 98GB-0012163.
XX PI (SMIK ) SMITHLINE BEECHAM BIOLOGICALS.
XX PI Ruelle J, Tommassen JPM, Vinals-Bassols C;
XX DR WPI; 2000-116523/10.
XX DR N-PSDB; AAZ29558.
XX KW Novel polypeptides used as vaccines for treating Maraxella catarrhalis
XX PT infections like otitis media and pneumonia
XX PS Claim 3; Page 102-103; 121pp; English.
XX CC The present sequence is the BASB031 polypeptide, from strain Mc2911,
XX CC derived from Moraxella catarrhalis. This sequence has homology to
XX CC Pseudomonas aeruginosa, PilQ fimbrial assembly protein. This sequence
XX CC can be used for prevention and treatment of M.catarrhalis infections,
XX CC like otitis media, pneumonia, sinusitis and nosocomial infections. The
XX CC antibodies and polynucleotide sequence can be used for diagnosing
XX CC infections, staging of disease and for determining the response of an
XX CC infectious organism to drugs. The DNA sequence is also used as a
XX CC hybridisation probe in screening process for identifying homologue and
XX CC orthologues from other species.
XX SQ Sequence 473 AA;
    Query Match 17.3%; Score 664.5; DB 21; Length 473;
    Best Local Similarity 34.4%; Pred. NO. 6.4e-41;
    Matches 159; Conservative 87; Mismatches 149; Indels 67; Gaps 11;
QY 339 APKFTTGRKISLDFQDVETIRILOILAKESGMNIVASDSVNGKWTLSLKDVDPDQALDLV 398
DB 38 AEQHYTGKPISEFADIPVRAVFDILAGFTGINITDSDSVTGMTIRLNPWDQAFDVI 97
QY 399 MQARNLDMRQOQGNIVNIAPRD-----ELLAKDRAFLQAEKDIAIDLALYSQNFQLK 449
DB 98 LQTNLSVLKHGVMWLISSKSIQSNQPTITEYIRLNYALAD---DVATL--IMGEKTQRG 152
QY 450 YKNVEFRSIL-----RLDNADTTGNRTLVSGRGSVLIDPATNLTIVTDTRS 497
DB 153 NVNRTNHRDVIYPEAPLSRAYNTQTDELITTAVRGTLTSSRGTVTDKRTNLTIIQDVP 212
QY 498 VIEKFRKILDELDPVPAQOVMEARIVEAADGFSRDLGVKFGATGKK-----K 544

```



Db 213 SVANIKALIERIDIPVEQVMEIARIIVSANENFGRKLGVSFGAHGQNGKHYGSGQSLWT 272  
Qy 545 LKNDTSFAGMGVNS-----GFGGDKWG--AETKINLPITAAANSISILVRAISSGALNLEL 598  
Db 273 MRQEGVAAGGHQSHLVNLDGVNMGRIAFGLNLP-----DIILDEL 317  
Qy 599 SASELSKTKTLANPRVLTQNRKEAKTESGVEIPFTVTSIANGSSNTTELKAVLGLTV 658  
Db 318 SAMQAEQGEVISTPKVLTADKQTARISSGLQIPYQETT--HSGAST-TKFEASLILEA 374  
Qy 659 TPNITPDGQIIMTVKINKDSPAQACASGNQTLICISTKNLNTQAMVENGGLIYGGIYEED 718  
Db 375 TPNITPDGKIGLKNIKNGNPVPTLGH---IAIQEDAETNVIIEGQVWLLGGIYRTS 430  
Qy 719 NGNLTQVPLGDIPTVGNLFKTRGKKTDRRELLIFITPRIM 760  
Db 431 QNEGANKVRLGDMFVLGRUFRHDSKAHDKSELLIFITPKLV 472

RESULT 9  
AA017573  
ID AAO17573 standard; Protein; 473 AA.  
XX AC AAO17573;  
XX DT 19-JUL-2002 (first entry)  
XX DE M catarrhalis MCA100992 protein SEQ ID NO: 26.  
XX KW Moraxella; vaccine; respiratory tract infection; antiinflammatory;  
XX KN auditory; antibacterial; otitis media; sinusitis; pneumonia.  
XX OS Moraxella catarrhalis.  
XX PN WO200218595-A2.  
XX PD 07-MAR-2002.  
XX PF 28-AUG-2001; 2001WO-CA01221.  
XX PR 28-AUG-2000; 2000US-228294P.  
XX PR 28-AUG-2000; 2000US-228295P.  
XX PR 28-AUG-2000; 2000US-228296P.  
XX PR 29-AUG-2000; 2000US-228438P.  
XX PR 29-AUG-2000; 2000US-228439P.  
XX PR 29-AUG-2000; 2000US-228440P.  
XX PR 29-AUG-2000; 2000US-228441P.  
XX PR 29-AUG-2000; 2000US-228442P.  
XX PR 29-AUG-2000; 2000US-228443P.  
XX PR 29-AUG-2000; 2000US-228511P.  
XX PR 29-AUG-2000; 2000US-228512P.  
XX PR 29-AUG-2000; 2000US-228742P.  
XX PR 29-AUG-2000; 2000US-228773P.  
XX PR 01-SEP-2000; 2000US-229465P.  
XX PR 01-SEP-2000; 2000US-229474P.  
XX PR 01-SEP-2000; 2000US-229475P.  
XX PR 01-SEP-2000; 2000US-229478P.  
XX PR 05-SEP-2000; 2000US-229740P.  
XX PR 05-SEP-2000; 2000US-229803P.  
XX PR 05-SEP-2000; 2000US-229804P.

Query Match 17.3%; Score 664.5; DB 21; Length 473;  
Best Local Similarity 34.4%; Pred. No. 6.4e-41;  
Matches 158; Conservative 87; Mismatches 153; Indels 61; Gaps 12;  
Qy 339 APKTFGTGRKISLDIPQDVEIRITLILAKESGMNIVASDSVNGKWTLSLKDVPHDQALDIV 398

Db 38 AEQHTGKPISEFADIPVRVDFILAGFTGNIITDSSVTGSMIRLNIWPDQAFDVI 97  
Qy 399 MOARLDMREQOQNIIVNIAPRD-----ELLAKDKAPLOAEKDIALDGAlySQNFOLK 449  
Db 98 LQTONLSVLKHCNWLSSKSIQSNQPTITEYIRLNYALAD---DVATL--IMGEXTORG 152  
Qy 450 YQNVFEFRSIL-----RLDNADTTGNRNTLVSGRGSVLIDPATNTLIVTDTRS 497  
Db 153 NYNRNHRDVIYPEAPLSAYNTQTDLELITTAARGTLJSERGTVIVDKRTNLIIDQVPA 212  
Qy 498 VIEKFKLIDELVPAQVQVMEIARIIVAEADGFSRDLGVKFGATGKKLKNDSATFQ---- 553  
Db 213 SVANIKALIERIDIPVEQVMEIARIIVSANENFGRKLGVSFGAHG---QNGKHYGSGQ 268  
Qy 554 -----WGVNSGFGGDDKWAETKINLPITAAANSISILVRAISSGALN-----LELSAS 601  
Db 269 SLWTMWEQVAAAGHQ--NSHLNVDLGVNMGRIAF-----GLNLPDIILDELSAM 320  
Qy 602 ELSLTKTKTLANPRVLTQNRKEAKIESGVEIPFTVTSIANGSSNTTELKAVLGLTVTPN 661  
Db 321 QAEQGEVISTPKVLTADKQTARISSGLQIPYQETT--HSGAST-TKFEASLILEATPN 377  
Qy 662 ITPDGOIMTVKINKDSPAQACASGNQTLICISTKNLNTQAMVENGGLIYGGIYEEDNGN 721  
Db 378 ITPDGKIGLKNIKNGNPVPTLGH---IAIQEDAETNVIIEGQVWLLGGIYRTSQNE 433  
Qy 722 TLTKVPLGDIPTVGNLFKTRGKKTDRRELLIFITPRIM 760  
Db 434 GANKVRLGDMFVLGRUFRHDSKAHDKSELLIFITPKLV 472

RESULT 10  
AA017573  
ID AAO17573 standard; Protein; 473 AA.  
XX AC AAO17573;  
XX DT 19-JUL-2002 (first entry)  
XX DE M catarrhalis MCA100992 protein SEQ ID NO: 26.  
XX KW Moraxella; vaccine; respiratory tract infection; antiinflammatory;  
XX KN auditory; antibacterial; otitis media; sinusitis; pneumonia.  
XX OS Moraxella catarrhalis.  
XX PN WO200218595-A2.  
XX PD 07-MAR-2002.  
XX PF 28-AUG-2001; 2001WO-CA01221.  
XX PR 28-AUG-2000; 2000US-228294P.  
XX PR 28-AUG-2000; 2000US-228295P.  
XX PR 28-AUG-2000; 2000US-228296P.  
XX PR 29-AUG-2000; 2000US-228438P.  
XX PR 29-AUG-2000; 2000US-228439P.  
XX PR 29-AUG-2000; 2000US-228440P.  
XX PR 29-AUG-2000; 2000US-228441P.  
XX PR 29-AUG-2000; 2000US-228442P.  
XX PR 29-AUG-2000; 2000US-228443P.  
XX PR 29-AUG-2000; 2000US-228511P.  
XX PR 29-AUG-2000; 2000US-228512P.  
XX PR 29-AUG-2000; 2000US-228742P.  
XX PR 29-AUG-2000; 2000US-228773P.  
XX PR 01-SEP-2000; 2000US-229465P.  
XX PR 01-SEP-2000; 2000US-229474P.  
XX PR 01-SEP-2000; 2000US-229475P.  
XX PR 01-SEP-2000; 2000US-229478P.  
XX PR 05-SEP-2000; 2000US-229740P.  
XX PR 05-SEP-2000; 2000US-229803P.  
XX PR 05-SEP-2000; 2000US-229804P.



Db 310 PQNILIRADESLNALVLLADPDVTATLEEIVRNLDVPRQAVMVEAAIVEISGDISDALGV 369  
 QY 536 KFGATGKKLKNDSYAFSGWVNSGFGG-----DDKWAETKINLPITAA 579  
 Db 370 QWAVDA-----RGGTGGLG-GVNFNTGLSVGTULKAQNEIIPD-----LTLP 413  
 QY 580 ANSILVRAISSGALNLELSASESLKTKTLANPRVLTQRKEAKIESGYEIPF---TVT 636  
 Db 414 DGAIIIGTGFALITALSAN---SKSNLLSTPSLLTLDNOEAEILVGNVPFQTGSYT 470  
 QY 637 SIANGSSTNTELKKAVALGLT--VTPNITPDGOIIMTV--KINKDSPAQASGNQTILCI 692  
 Db 471 TDAGANNPFTTIEREDIGVTLKVTPHINDGATLRLEVEQEISSIAPSAGVNAQAVDLVT 530  
 QY 693 STKNLNTQAMVNGGTLIVGGIYEEDNGNTLTKVPLLDGIPVIGNLFKTRGKKTDRRELL 752  
 Db 531 NKRSIKSVILADDDGVVLGGLIQDDVTSTDSKVPILGDIPLIGLFRSTKDTHTVKNLNM 590  
 QY 753 IFITPRIM 760  
 Db 591 VFLRPTIV 598

## RESULT 12

AA82594 ID AA82594 standard; Protein; 649 AA.

AC AA82594;

XX 01-AUG-2000 (first entry)

XX Pseudomonas alcaligenes XcpQ protein sequence SEQ ID NO:15.

XX Pseudomonas alcaligenes; expression; lipase regulation cascade;  
 KW kinase; DNA binding regulator; polymerase; promoter; secretion factor;  
 KW XcpP; XcpQ; XcpR; XcpS; XcpT; XcpU; XcpV; XcpW; XcpX; XcpY; XcpZ;  
 KW OrfY; OrfZ; OrfZ; LipQ; LipR; upstream activating sequence;  
 KW detergent; cleaning formulation.

XX Pseudomonas alcaligenes.

XX US6048710-A.

XX 11-APR-2000.

XX 15-AUG-1997; 97US-0911853.

XX 16-AUG-1996; 96US-0699092.

XX (GEMY) GENENCOR INT INC.

XX Gerritse G, Quax WJ;

XX WPI; 2000-316896/27.

XX N-PSDB; AAA13897.

XX Expression vector for producing heterologous proteins in host cells  
 PT comprises a nucleic acid encoding a kinase and a DNA binding regulator  
 PT which hybridizes under stringent conditions to nucleic acid

XX Example 7; Fig 3; 133pp; English.

XX The present invention describes an expression vector comprising a  
 CC nucleic acid encoding a kinase and a DNA binding regulator which  
 CC hybridises under stringent conditions to a nucleic acid isolated  
 CC from Pseudomonas alcaligenes. Also described are: (1) an isolated  
 CC plasmid comprising the above expression vector; (2) a method of  
 CC transforming a host cell comprising adding the above plasmid to host  
 CC cells under appropriate conditions; (3) a transformed host cell  
 CC comprising the above expression vector; and (4) a method for producing  
 CC a protein comprising the steps of obtaining a host cell comprising the  
 CC above expression vector and further comprising nucleic acid encoding  
 CC the protein, and culturing the host cell under conditions for the

CC expression of protein. The expression vector of the present invention  
 CC can be used for producing heterologous proteins in host cells,  
 CC particularly, lipase in Pseudomonas. Lipases produced can be used in  
 CC detergents and cleaning formulations in industrial processes. The  
 CC invention provides a higher production level and efficiently express a  
 CC heterologous protein. The present sequence represents XcpQ isolated from  
 CC Pseudomonas alcaligenes, from the present invention.

XX Sequence 649 AA;

Query Match 8.9%; Score 341.5; DB 21; Length 649;  
 Best Local Similarity 22.4%; Pred. No. 1.2e-16;  
 Matches 136; Conservative 94; Mismatches 183; Indels 195; Gaps 18;

QY 315 AAFGYFTFVLPRKKNLESGGVNNAKPTFTGRKISLDFQDFQVIRITLQILAKESGMNIVA 374  
 Db 24 AAP-----LPLVHAAPVAVSQGAETWT-----INMKDADIRDFDQVAISGETFVV 71

QY 375 SDSVNGKMT-----LSLKDVPWDQALDVMQARNLDMRQQGNVYNIAPRELLAKDKAF 428  
 Db 72 DPRVKGQVTIVISKTPGLLEEV--YQLFLSVNTHGFSVLAQGDQARIVPVTARSANSS 129

QY 429 LOAEKDI-----ADLGALYSQNFQ---KYKNVEEFRSI 459  
 Db 130 RSAPDDVQTELIOVQHTSVNELIPLIRPLVPQNGHLAAVAASNALIISORRANIEREL 189

QY 460 LR-----LDNADTTGN---RNTLVSGRGSVLIDPATNTLIV---- 492  
 Db 190 IAEFDAQGGGDYNNVINLOHAWVLDAEALNNVAMRNEKNSAGTRVIADARTNRLILLGPP 249

QY 493 -----TQTRSVIE----- 500  
 Db 250 AAFQRLANLARSIDIPSTRSANARVILRHSDAKSLAETLGDISEGLKTAEGGEAASSK 309

QY 501 -----KFKLIDELVPAQOVMIARIVAAAGFSRDLGV 535  
 Db 310 PQNILIRADESLNALVLLADPDVTATLEEIVRNLDVPRQAVMVEAAIVEISGDISDALGV 369

QY 536 KFCATCKKKLXNDTSAPFGWVNSGFGG-----DDKWAETKINLPITAA 579  
 Db 370 QWAVDA-----RGGTGGLG-GVNFNTGLSVGTULKAQNEIIPD-----LTLP 413

QY 580 ANSILVRAISSGALNLELSASESLKTKTLANPRVLTQRKEAKIESGYEIPF---TVT 636  
 Db 414 DGAIIIGTGFALITALSAN---SKSNLLSTPSLLTLDNOEAEILVGNVPFQTGSYT 470

QY 637 SIANGSSTNTELKKAVALGLT--VTPNITPDGOIIMTV--KINKDSPAQASGNQTILCI 692  
 Db 471 TDAGANNPFTTIEREDIGVTLKVTPHINDGATLRLEVEQEISSIAPSAGVNAQAVDLVT 530

QY 693 STKNLNTQAMVNGGTLIVGGIYEEDNGNTLTKVPLLDGIPVIGNLFKTRGKKTDRRELL 752  
 Db 531 NKRSIKSVILADDDGVVLGGLIQDDVTSTDSKVPILGDIPLIGLFRSTKDTHTVKNLNM 590

QY 753 IFITPRIM 760

Db 591 VFLRPTIV 598

## RESULT 13

AA82256

ID AA82256 standard; Protein; 649 AA.

XX AA82256;

XX 09-JUL-2001 (first entry)

XX Pseudomonas alcaligenes secretion factor XcpQ.

XX Lipase; XcpQ; secretion factor; protein secretion.

XX Pseudomonas alcaligenes.

PN US6225106-B1.  
 XX  
 PD 01-MAY-2001.  
 XX  
 PF 06-JAN-2000; 2000US-0479409.  
 XX  
 PR 15-AUG-1997; 97US-0911853.  
 PR 16-AUG-1996; 96US-0699092.  
 XX  
 PA (GEMV ) GENENCOR INT INC.  
 XX  
 PI Gerritse G, Quax WJ;  
 XX  
 DR WPI; 2001-315684/33.  
 DR N-PSDB; AAF30870.  
 DR AAB82262, AAB82263, AAB82264, AAB82265, AAB82266, AAB82267,  
 DR AAB82268, AAB82269.  
 XX  
 Novel isolated nucleic acid encoding kinase from *Pseudomonas* that can  
 PT regulate expression of lipase, useful in expression systems for  
 PT production of lipase which is useful in detergents and cleaning  
 PT formulations .  
 XX  
 PS Example 7; Fig 3A; 131pp; English.  
 XX  
 CC The present sequence is that of secretion factor XcpQ of  
 CC *Pseudomonas* alcaligenes. The secretion factor is encoded by an  
 CC open reading frame identified in a cosmid (see AAF30870) derived  
 CC from *P. alcaligenes* DNA. Secretion factors aid the secretion of  
 CC other proteins from a cell. A new expression system comprises  
 CC components of a lipase regulation cascade including a kinase, DNA  
 CC binding regulator, polymerase, a promoter, an upstream activating  
 CC sequence, and secretion factors. The secretion factor is preferably  
 CC a member of the Xcp protein family (see AAB82256-68) and acts in  
 CC concert with other members of the Xcp family. Plasmids and  
 CC transformed cells are provided, and also host cells which further  
 CC comprise a nucleic acid encoding a desired protein, especially an  
 CC esterase, hydrolase, lipase, isomerase, mutase, transferase, kinase  
 CC or phosphatase (claimed). A hyper-producing strain can be  
 CC developed.  
 XX  
 SQ Sequence 649 AA;  
 Query Match 8.9%; Score 341.5; DB 22; Length 649;  
 Best Local Similarity 22.4%; Pred. No. 1.2e-16;  
 Matches 136; Conservative 94; Mismatches 183; Indels 195; Gaps 18;  
 QY 315 AAPGYFTFOVLPKKNLESQGVNNAKFTGCKISLDFQDVEIRTILOILAKESGMNIVA 374  
 DB 24 AAP-----LPLVHAAEPVAVSQGAETWT-----INMKDADIRDFIDQVAQISGETFW 71  
 QY 375 SDSVNGKMT-----LSLKDVDPDQALDVMQARNLDMRQQGNIVNIAPRDELLAKDKAF 428  
 DB 72 DPRVKGQTVISKTPILGLEEV--YQLFLVMSTHGFSLAQDQARIIVPVTARSGANSS 129  
 QY 429 LQAEKDI-----ADLGALYSONFOL-----KYNVEEFERSI 459  
 DB 130 RSAPDDVQTELIQVQHTSVNELIPIRPIVLPQNGHLAAVAASNALIISDRANIERREL 189  
 QY 460 LR-----LDNADTGN---RNTLVSRGSLVIDPANTLIV----- 492  
 DB 190 IAEIDAQGGDYVINLQHAWLDAEALNNAVMNKNKSAGTRVIADARTNRLILGPP 249  
 QY 493 -----TTTRSVIE----- 500  
 DB 250 AARQRLANLARSLDIPSTRSANARVIRLHSDAKSLAETLGISGLKTARGGGEAASSK 309  
 QY 501 -----KFKXILDELDPQAQVMIEARIIEAADGFSRDLGV 535  
 DB 310 PQNILLIRADESLNALVLLADPTVATLEIVNLNDVPRAQVMVEAAIVEISGDISDALGV 369  
 QY 536 KFGATGKKXKKNKDTSAFGWNGVSGFG-----DDKWAETKINILPITAA 579

DB 370 QWAVDA---RGGTGGLG-GVNFNGTGLSVGTVLKAIQNEEIPDD-----LTLF 413  
 QY 580 ANSISLVRAISSGALNLELSASELSKTKTLANPRVLTONTREAKIESGYEIPF---TVT 636  
 DB 414 DGAITGIGTENFQALITALSAN---SKNLLSTPSELLTLDNQEAELLVQNVYFQGSYT 470  
 QY 637 SIANGGSSTNTELKXAVGLT--VTPNITPDGQIIMTV--KINKSPACASGNQITLCI 692  
 DB 471 TDASGANPFETIEREDIGVTLKVTPIHINDGATLARLEVEQEISSIAPSAGVNAQAVDLVT 530  
 QY 693 STKNLNTQAMVENGGLIIVGGIYEEDNGNTLTKVPLLDGIPVIGNLFTKRGKTKDRRELL 752  
 DB 531 NKRSIKSVILADDDQVIVLGGLIQDDVTSTDSKVPILGDIPLIGLFRSTKOTHVKNLML 590  
 QY 753 IFITPRIM 760  
 DB 591 VFURPTIV 598  
 RESULT 14  
 AAE13660  
 ID AAE13660 standard; Protein; 649 AA.  
 XX  
 AC AAE13660;  
 XX  
 DT 26-FEB-2002 (first entry)  
 XX  
 DE *Pseudomonas* alcaligenes XcpQ secretion factor protein.  
 XX  
 KW DNA binding regulator; lipase regulation cascade; kinase; polymerase;  
 KW promoter; upstream activating factor; secretion factor; detergent;  
 KW cleaning formulation; industrial process; XcpQ protein.  
 XX  
 OS *Pseudomonas* alcaligenes.  
 XX  
 PN US6313283-B1.  
 XX  
 PD 06-NOV-2001.  
 XX  
 PF 06-JAN-2000; 2000US-0479453.  
 XX  
 PR 15-AUG-1997; 97US-0911853.  
 PR 16-AUG-1996; 96US-0699092.  
 XX  
 FA (GEMV ) GENENCOR INT INC.  
 XX  
 PI Gerritse G, Quax WJ;  
 XX  
 DR WPI; 2002-024912/03.  
 DR N-PSDB; AAD22874.  
 XX  
 PT Polynucleotide encoding DNA binding regulator, useful for regulating  
 PT expression of lipase especially in *Pseudomonas*, and lipase useful in  
 PT detergents and other cleaning formulations and in various industrial  
 PT processes .  
 XX  
 PS Disclosure; Column 37-40; 132pp; English.  
 XX  
 CC The present invention relates to a polynucleotide encoding a DNA  
 CC binding regulator that can regulate the expression of a lipase. The  
 CC invention also relates to an expression system comprising components  
 CC of *Pseudomonas* alcaligenes lipase regulation cascade which includes  
 CC kinases, DNA binding regulators, polymerases, promoters, upstream  
 CC activating factors and secretion factors. DNA binding regulators  
 CC of the invention are useful for regulating the expression of a  
 CC lipase, where the lipase is useful in detergents and other cleaning  
 CC formulations as well as a number of industrial processes. Various  
 CC components of lipase regulation cascade is useful in expression  
 CC methods and systems designed for the production of heterologous  
 CC proteins. The present sequence is *P. alcaligenes* XcpQ, a secretion  
 CC factor protein.  
 XX  
 SQ Sequence 649 AA;



QY 613 PRVLTONRKEAKIESGYEIPFTVTSIANGSSSTNTEKKAVALG--LVTNPTNITPDGQIIM 670  
 DB 450 PSIVTLDNKLASFNVGDVPLVSGSQTTSGDNVNTVERKTGKLVTPQVN-EGDAVL 508  
 QY 671 TVKINKDSPAQCASGNQTI-LCISTKNLNTQAMVENGSTLIVGIYEDNGNTLTKVPLL 729  
 DB 509 -LEIBQEVSSVDSSENSTLGTFTNTRITQNAVLKVTGETVVLGGLDDFSKEQVSKVPLL 567  
 QY 730 GDIPVIGNLFKTRGKKTDRRELLIFITPRIM 760  
 DB 568 GDIPVIGQLFRYSTERAKRNLVFIPTII 598

## RESULT 16

AAU33596  
 ID AAU33596 standard; Protein; 776 AA.

AC AAU33596;  
 XX  
 DT 14-FEB-2002 (first entry)  
 DE Pseudomonas aeruginosa cellular proliferation protein #40.  
 KW Antisense; prokaryotic cellular proliferation protein;  
 KW antibiotic; antibacterial; drug design.  
 XX  
 OS Pseudomonas aeruginosa.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 23-OCT-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI: 2001-611495/70.

DR N-PSDB; AAS51455.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 5092; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes, their use in the discovery of novel antibiotics, the essential

CC genes themselves and the encoded proteins. The prokaryotes used are

CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*

CC *aeruginosa*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The

CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used

CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen

CC for homologous nucleic acids which are required for cell proliferation in

CC a wide variety of organisms. The present sequence represents an

CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC

CC

CC

CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 776 AA;

QY Query Match 7.8%; Score 300; DB 22; Length 776;

Best Local Similarity 21.4%; Pred. No. 1.9e-13;

Matches 122; Conservative 97; Mismatches 180; Indels 170; Gaps 16;

QY 349 SLDFQDVEIRITLILAKESGMNIVASDSVNGKMT-----LSLKDVDPWDQALDVMQAR 402

DB 47 TINMKDAEIGDFTIEQVSSISGQTFVVDPRVKGRVTWVSOARLSLAEV--YQLFSLVLATH 104

QY 403 NLDMRQGNIVNIAPRDELLAKDKAFLOAEKIAD----- 437

DB 105 GYAVLPQGDQARIVPNME--ARQDA---AOKTVRDGPGSLETRVVOAQOTSVAELIPMR 159

QY 438 -----LICALYSONFQL---KYKNVEEFRSILR----- 461

DB 160 PLVPAHGHLAAVPSANALIVSDRRSNIERIEAIVRSIDRAGEHDYSIYDMRHAWVAEIAE 219

QY 462 -LDN-----AOTTGNRTLV----- 475

DB 220 VLDRSVTPAAGKSAATVQLADSRNRLVLLGPPQARAEALLRLAQSLDVPSSRSANSRVI 279

QY 476 -----SGRGS-----VLIDPATNTLIVTDTRSVIEKF 502

DB 280 RLRHGDAKTLAATLGEISLHGERGQDGRGKGLLVRADESINALVILADPEDVGLL 339

QY 503 RKLIDELDVPAAQVMIEAIVEAAGDSRDGLVKFGATGKKLKNDSAFGNG-VNSGFG 561

DB 340 EDIVRQLDVPRAQLLVEAAIVELSGEIGALGVQWA-----LRSGHVAGGAGFADSGLS 393

QY 562 GDSKGAETKINLPITAAANSISLVRAISSGALNLSASELSKTKTLANPRVLTONRK 621

DB 394 IGTLLGALQAGKPPAELPDCAIVGLSRDFGALVTALSRN---SRNLSLSTPSLLTLDNQ 450

QY 622 EAKIESGYEIPP---TVTSTANGSSSTNTEKKAVALGLT--VTPNITPDGQIIMTV--KI 674

DB 451 KAEILVQNVFPQTGSGVTTASAGSSNPFTTVERKDIGVTLKVTPHIGEDRMRLAEIEQEI 510

QY 675 NKDSPAQCASGNQTIILCISTKNLNTQAMVENGSTLIVGIYEDNGNTLTKVPLLGDIPV 734

DB 511 SSIAPTATLAAKAVDLVTNRKRSIKSTVLADGGQVIVLGLIQDLDORSRVELLDGIPG 570

QY 735 IGLNFKTRGKKTDRRELLIFITPRIMGTA 763

DB 571 VGLFRSSRRETRVKRNLVFLRPSIVRDA 599

RESULT 17

AAU37501

ID AAU37501 standard; Protein; 666 AA.

XX AC AAU37501;

XX DT 07-OCT-1999 (first entry)

XX Chlamydia trachomatis surface exposed protein.

CC Vaccine; eye disease; conventional trachoma; nonendemic trachoma;

CC paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;

CC nongonococcal urethritis; epididymitis; cervicitis; salpingitis;

CC Bartholinitis; pneumopathy; venereal lymphogranulomatosis.

XX Chlamydia trachomatis.

XX WO9928475-A2.

XX 10-JUN-1999.

XX 27-NOV-1998; 98WO-IB01939.

XX 04-NOV-1998; 98US-0107077.

PR

|    |                      |               |  |
|----|----------------------|---------------|--|
| PR | 28-NOV-1997;         | 97FR-0015041. |  |
| PR | 17-DEC-1997;         | 97FR-0016034. |  |
| XX |                      |               |  |
| XX | (GEST ) GENSET.      |               |  |
| XX | Griffais R;          |               |  |
| XX |                      |               |  |
| XX | WPI; 1999-371125/31. |               |  |
| DR |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |
| XX |                      |               |  |

Db 519 ERLFLRPRWAIEG 533

RESULT 19  
AAY35326  
ID AAY35326 standard; Protein; 561 AA.  
XX AC AAY35326;  
AC XX  
DT 13-SEP-1999 (first entry)  
XX Chlamydia pneumoniae surface exposed polypeptide.  
DE Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope.  
XX Chlamydia pneumoniae.  
OS OS  
XX WO9927105-A2.  
PN PD  
PD 03-JUN-1999.  
XX 20-NOV-1998; 98WO-IB01990.  
PF XX  
XX 04-NOV-1998; 98US-0107078.  
PR 21-NOV-1997; 97FR-0014673.  
PA (GEST ) GENSET.  
XX Griffais R;  
PI WPI; 1999-357842/30.  
DR XX  
XX Genome sequence of Chlamydia pneumoniae  
PT Page 1134-1135; Disclosure; 1912pp; English.

XX AAY34584-Y35879 represent the proteins encoded by all the open reading  
frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.  
CC C. pneumoniae causes respiratory disease such as pneumonia and  
CC bronchitis and is thought to be a contributing factor in heart  
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
CC nucleotide sequences can also be used as immunogenic compositions,  
CC especially where the vector directs the expression of a neutralising  
CC epitope of C. pneumoniae.  
XX Sequence 561 AA;

Query Match 6.1%; Score 233; DB 20; Length 561;  
Best Local Similarity 22.3%; Pred. No. 1.1e-08;  
Matches 106; Conservative 91; Mismatches 183; Indels 96; Gaps 17;

QY 349 SLDFQDVEIR-----TILQLAKES-----GMN---IVASDVNGKMTLS 385  
DB 66 SVDMEYEVKYANPALVSVCQDVLGTLAEDDAFCQMFTOPGNKIFVSSPRLANKAEQL 125  
QY 386 LKVDPDQDLVMQARNLDMRGQNVNIAPRELLAKDKAFLOAEKDIALGLAYSQN 445  
DB 126 LK-----SLDVPMAHTLD-----DPASTALAGGGTGTSFK-----SLRFFM 163  
QY 446 FOLKYNVEEFSRILR---LDNADTTGNRTNLVSGRGSVLIDPATNTLVITDTRSVIEKF 502  
DB 164 YKLKYQNEVIANALODIGLYNVLTAMDEDEINTLNLSIQWLEVNNSIVIIGNQGVDRV 223  
QY 503 RKLIIDEVPAQVMTAEARIVEAADGFSDLGKVGATCKKKLKNDTSAFGWG--NSGF 560  
DB 224 IGLNGLDLPPKQVYEVLIDTSLSEKSWDFGQVVALGDEQS---VAYASGLLNNTGI 280

QY 561 GGDDKMGAEETKINLP-----ITAAANSISILVRAISSG----- 591  
DB 281 -----ATPTKATVPFPGTPNPGSIPLPTEGQLTFGSDMLNSSSAFGUGIIGNVLSHKGKS 334  
QY 593 --ALNLELSASELSKTKTLANPRVLTONRKEAKIESGYEIPP--TVTSIANGSGST-NT 647  
DB 335 FLTLGGLLSALDQGDGTIVLNPRIQAQDTQQAGFFVGQTVPYQTTNTIIOETGTVTQNI 394  
QY 648 ELKKAUHLGTYTPNITPDGOIIMTVKINKD-SPAQCASGNOTILCISTKNLNTQAMVENG 706  
DB 395 DYEDIGVNLVTVTAPNN--VVTLQIEQTISELHSASGSLTPVTDKT-YAATRLOIPDG 451  
QY 707 GTLIVGGIYEEDNGNTLTKVPLLDGPVIGNLFKTRGKTKDRRELLIFITPRIMG 762  
DB 452 CFLVMSGHIRDKTTKWSGVPLLSIPLIRGLFSRTIDQROKRNMIMPFIKPVKISS 507

RESULT 20  
AAY37640  
ID AAY37640 standard; Protein; 701 AA.  
XX AC AAY37640;  
AC XX  
DT 07-OCT-1999 (first entry)  
XX Chlamydia trachomatis secreted protein.  
DE Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
XX paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;  
KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.  
XX Chlamydia trachomatis.  
OS OS  
XX WO9928475-A2.  
PN PD  
PD 10-JUN-1999.  
XX 27-NOV-1998; 98WO-IB01939.  
PF XX  
XX 04-NOV-1998; 98US-0107077.  
PR 28-NOV-1997; 97FR-0015041.  
PR 17-DEC-1997; 97FR-0016034.  
XX (GEST ) GENSET.  
PA Griffais R;  
XX WPI; 1999-371125/31.  
DR Genome sequence of Chlamydia trachomatis  
PT Disclosure; Page 1277; 1755pp; English.  
PS CC  
CC AAY3754-Y37949 are encoded by open reading frames (ORFs) of the genome  
of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as  
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
CC can also be used to control growth of the microorganism. Chlamydia  
CC trachomatis is responsible for a large number of diseases, e.g. eye  
CC diseases such as conventional trachoma, nonendemic trachoma, such as  
CC paratrachoma, and inclusion conjunctivitis; genital diseases, such as  
CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
CC perihepatitis, bartholinitis; pneumopathy in breast feeding infants;  
CC and venereal lymphogranulomatosis. The polypeptides of the invention  
CC may be of use in treating these diseases.  
XX Sequence 701 AA;

Query Match 6.0%; Score 229.5; DB 20; Length 701;  
Best Local Similarity 22.6%; Pred. No. 2.9e-08;  
Matches 130; Conservative 86; Mismatches 229; Indels 139; Gaps 23;

QY 241 ELAALGFAG---OPDISOOHDHIIIVTLKNHTLPTLQRSLSVDAPKFTPVQKVTLRKLNND 297



Db 138 EGVLFIFGSRQELSLPMTAHIAFLSSKNL-----DARADVQALR-KFANSQ 185  
Qy 298 TQLIIITAGNWLKNSAAGPYFTFQVLPKONLESGGVNNAFKTFTGRKISLDFQDVEI 357  
Db 186 TMLIDFICKVWLF---GAVSEIT-ELLKIYEFLOSHNIEEHXIVLSKI-----EPSM 237  
Qy 358 RTILOI-----LAKES-----GNIVASDSVNGKMTLSLKDVWQDALDLVWQARNLM 406  
Db 238 LXILKAARFREDLAKESGSGVGLKVPLQNHGRSLFLS-----GALPIVQKADILIR 290  
Qy 407 RQCGNIVNIAPRELLAKDAKFLQAEKDIALDLGALYSQNFQKLYKXVEEFSRILRLDNAD 466  
Db 291 ELEEGIE--SPTKTVP---WYHVKSDPOELALLSQVHDI-FSNGASCAS-----SSCD 340  
Qy 467 T-----TGNRNLVSGR---GSVLIDPATNTLIVTDRSVIEKFRK 504  
Db 341 TGVVSSKAGSSNGLGVHIDTSLRSSVKEGSAKYGSPFIADSKTGTILMIVIEKALPKIM 400  
Qy 505 LIDELDPVPAQVNIARIVEAADGFRDLGVKFGATGCKKLNKNDTSAFGNGVNSFGGDD 564  
Db 401 LKXKLOVPKXWRIEVLFPKLSNQRKSGNLRLRLOEEVCKGCTQAVSW-----450  
Qy 565 KWAETKINLPITAAANSISLVRAISGALNLE-----LSASESLSKTKTLANPRV 615  
Db 451 -----ASGGILEFLFKGAKGIVPSYDFAYQFLMAQEDV---RINASPSV 492  
Qy 616 LTQNRKEAKIESGYEIPFTVTSTANGSSNTTELKXAVLGLTVTPNITP-----DGQI 668  
Db 493 VTWNTQPARIAIWEWSIVVSS-----DKDKAQINRAQYIMI--KILPVINIGEEDGKS 545  
Qy 669 IMTVK--INKDSFAQCASGNQTLICSTKNLNTQAMVENGGLIVGGIYEEDNGNLTIV 726  
Db 546 FITLETDTITDSTGRNHADRPD---VTRRNITNKVRIQDGETVIIGLRNQNMTDSRDI 602  
Qy 727 PLIGDIPVIGNLPKTRCKKTDRELLIFITPRIM 760  
Db 603 PFLGELPGICKLFGMSASDSQTEMENFITPKIL 636  
RESULT 21  
AAV92827  
ID AAV92827 standard; Protein; 698 AA.  
AC AAV92827;  
XX 29-AUG-2000 (first entry)  
DT 29-AUG-2000 (first entry)  
XX C. pneumoniae CPN100538 antigen.  
XX Antigen; anti-inflammatory; respiratory; antibacterial; anti-asthmatic;  
KW anti-arteriosclerotic; vaccine.  
XX Chlamydia pneumoniae.  
OS W0200024765-A2.  
PN 04-MAY-2000.  
PD 28-OCT-1999; 99WO-CR00992.  
PF 28-OCT-1998; 98US-0106034.  
PR 28-OCT-1998; 98US-0106039.  
PR 28-OCT-1998; 98US-0106042.  
PR 28-OCT-1998; 98US-0106044.  
PR 28-OCT-1998; 98US-0106072.  
PR 28-OCT-1998; 98US-0106073.  
PR 28-OCT-1998; 98US-0106074.  
PR 28-OCT-1998; 98US-0106087.  
PR 02-NOV-1998; 98US-0106587.  
PR 02-NOV-1998; 98US-0106588.  
PR 02-NOV-1998; 98US-0107034.  
PR 02-NOV-1998; 98US-0107035.

XX PA (CONN-) CONNAUGHT LAB LTD.  
XX PI Murdin AD, Oomen RP, Wang J;  
XX DR WPI: 2000-350688/30.  
XX N-PSDB; AAA28704, AAA28705.  
XX Chlamydia antigens and the proteins they encode, useful for  
PT vaccinating against Chlamydia infections that affect the respiratory  
PT tract  
XX Claim 13; Fig 15; 226pp; English.  
XX The nucleic acids may be used for the recombinant production of the  
CC Chlamydia polypeptides (either in vivo or in vitro) according to standard  
CC recombinant DNA methodologies. The polypeptides may then be used to  
CC vaccinate against Chlamydia infections in mammals. Chlamydia, such as  
CC C. pneumoniae, are pathogens responsible for upper respiratory tract  
CC infections such as community acquired pneumonia, acute respiratory  
CC disease and bronchitis and may be implicated in atherosclerotic changes  
CC and asthma. The nucleic acids may also be used as probes for detecting  
CC the presence of Chlamydia nucleic acids in samples (and therefore  
CC diagnose infections) and the proteins may be used as antigens for the  
CC production of antibodies that may be used to detect Chlamydia proteins  
CC in samples (e.g. via enzyme linked immunosorbant assay (ELISA)).  
XX SQ Sequence 698 AA;  
Query Match 5.9%; Score 226; DB 21; Length 698;  
Best Local Similarity 21.8%; Pred. No. 5.2e-08;  
Matches 104; Conservative 91; Mismatches 185; Indels 96; Gaps 17;  
Qy 349 SLDFODVIR-----TILQILAKES-----GNV-----IVASDSVNGKMTLS 385  
Db 203 SVDMTEYEVKANPAALVSYQDVLTGLTAEDDAFQFOTGTNKFVSVSSRLANKAQL 262  
Qy 386 LKDVPMQDALVMOARNLDMRQGNIVNIAPRDELLAKDAKFLQAEKDIALDLGALYSQN 445  
Db 263 LK-----SLDVPENAHITL-----DPASTALAGGTGTSPP-----SLRPFM 300  
Qy 446 FOLKYKVEEERSILR---LONADTGNRNLVSGRGSVLIDPATNTLIVTDRSVIEKF 502  
P 301 YKLKYQNGEVITANALQDIGNLYVTAMDDEFINTLSIQMLEVNSIVITIGNQGNVDRV 360  
Qy 503 RKLIDELVPAQVNIARIVEAADGFRDLGVKFGATGCKKLNKNDTSAFGNGV--NSGF 560  
Db 361 IGLNGLDLPPKQVYIEVLIDTSLKSWDFGVQWVALGDSQSK---VAYASGLLNNTGI 417  
Qy 561 GGDCKWGAETKINLP-----ITAAANSISLVRAISSG-----592  
Db 418 -----ATPTKATVPPTGPNFSGISPLPTPGQLTGFSDMLNSSAFGLGIQNVLSHKGS 471  
Qy 593 --ALNLELSASESLSKTKTLANPRVLNTQNRKEAKIESGYEIPFTVTS--IANGGSST-NT 647  
Db 472 FLTGLGILLSALDDQDGTIVILNPRIMAQDTQQAFFVGQTVPYQTIKYIQTGTVTQNI 531  
Qy 648 ELKKAVLGLTVTPNITPDGQIIMTVKINKD-SPAQCASGNQTLICSTKNLNTQAMVENG 706  
Db 532 DYEDIGVNLVVTSTVAPNN--VVTLLQIEQTISELSASGSLTPVTDKT--YAATRLQIPDG 588  
Qy 707 GTLIVGGIYEEDNGNLTQVPLLGIPVIGNLFKTRGKKTDRRELLIFITPRIMGT 762  
Db 589 CFLVMSGHIRKTKTVKSGVPLLSIPLIRGLFSTRIDORQKXNIMFIKRVISS 644  
RESULT 22  
AAV35456  
ID AAV35456 standard; Protein; 754 AA.  
XX AAV35456;  
XX AC  
XX 13-SEP-1999 (first entry)  
DT

XX DE Chlamydia pneumoniae transmembrane protein sequence.  
XX KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope.  
XX OS Chlamydia pneumoniae.  
XX PN WO9927105-A2.  
XX PD 03-JUN-1999.  
XX PF 20-NOV-1998; 98WO-IB01890.  
XX PR 04-NOV-1998; 98US-0107078.  
XX PR 21-NOV-1997; 97FR-0014673.  
XX (GEST ) GENSET.  
XX Griffais R;  
XX WPI; 1999-357842/30.  
XX Genome sequence of Chlamydia pneumoniae  
XX Page 1229-1230; Disclosure; 1912pp; English.  
XX AAY34584-Y35879 represent the proteins encoded by all the open reading  
frames in the complete genome (see AAY91990) of Chlamydia pneumoniae.  
XX C. pneumoniae causes respiratory disease such as pneumonia and  
XX bronchitis and is thought to be a contributing factor in heart  
XX disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
XX nodosum or pharyngitis. The polypeptides encoded by the open reading  
frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
XX immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
XX nucleotide sequences can also be used as immunogenic compositions,  
XX especially where the vector directs the expression of a neutralising  
XX epitope of C. pneumoniae.  
XX SQ Sequence 754 AA;  
Query Match 5.5%; Score 212.5; DB 20; Length 754;  
Best Local Similarity 21.3%; Pred. No. 6e-07;  
Matches 111; Conservative 82; Mismatches 217; Indels 111; Gaps 16;  
287 QKYLKRLNNDQLIITAGNHWLVNKSAPGYFTFQVLVLPKQNLSEGGVNNAPKTFGR 346  
234 QHVLKFFINPETHVDVIAGRVNIFGSAEYGV-----ELKIYNFVQSSIRQYRVIPLT 289  
347 KISLDPQDVEIRITLQILAKESGMNIVASDVNGKMTLSLKDVPWDQ-----ALDL 397  
290 KI-----DPGEMISILNAAFREDLTKDVSSES-----LGLRVVPLQYQGRSLFLSGTAAL 339  
398 VVQARNLDNRQGNINVIAPRELLAKDKAFQAEKDADLQALYQNFQLYKYNVEFR 457  
340 VQQAALTILRELEGIEIN--PTDKTVF---WYNVXHSDDPQELAAALSQ-----VHDFVSGE 389  
458 SILRLNDATNGR-----NTLVSGR-----GSLIDPANTLIVTDTSRVIEKF 502  
390 NKASVGAADGCGSQLNASCIDTIVSSAKDGSVKYGNFIADSKTGLIMVVEKEVLPRI 449  
503 RKUIDELVPAQOQVMEARIVEAADGFSRDLGVFGATGKKLKNDDTAFAGVNGVSGFG 562  
450 QMLLKKLDVFKMVRVIEVLFFERKLAHEQKQSGNLRLRGEYVCKKGS----- 497  
563 DRKWAETKINLPITAAANSISLVRAISSGALNLELSASESL-----KTKTLANPR 614  
498 -----PSVSWAGGTGILELFLKSGTSSIVPGYDLAYQFLMAQEDVRINASPS 545  
615 VLTQNRKAKIESGYIPTVTSIANGSGSTWTELKAVILGTV--TPNIT---PDGQII 669  
546 VVTMNTQTPARIAYVDEM-----SIAYVSDKDKAQYNAQYGMIRKMLPVINVGEEGKSY 600

QY 670 MTV-----KINKSPAQCASGNQITLCISKNLNTQAMVNGGTLIVGIVEEDN 719  
DB 601 ILETDITPTDKXDDRPD-----VTRNITNKVRIADGETVILGLRCKQM 649  
QY 720 GNTLTKVPLLGIPVIGNLFKTRGKKTDRRELLIITPRIM 760  
DB 650 SDSDHGIPFLGDIPGIGIKLFGMSSTSDSLTEMFVITPKIL 690  
RESULT 23  
ABB90607  
ID ABB90607 standard; Protein; 754 AA.  
XX AC ABB90607;  
XX 29-JUL-2002 (first entry)  
XX Chlamydia pneumoniae cp7127 protein, SEQ ID NO:163.  
XX Chlamydia pneumoniae.  
XX Key Location/Qualifiers  
FT Peptide 1..23  
FT Protein /label= Signal\_peptide  
FT Protein 24..754  
FT Protein /note= "Mature protein"  
XX WO200202606-A2.  
XX 10-JAN-2002.  
XX 03-JUL-2001; 2001WO-IB01445.  
XX 03-JUL-2000; 2000GB-0016363.  
XX 11-JUL-2000; 2000GB-0017047.  
XX 21-JUL-2000; 2000GB-0017983.  
XX 07-AUG-2000; 2000GB-0019368.  
XX 18-AUG-2000; 2000GB-0020440.  
XX 14-SEP-2000; 2000GB-0022583.  
XX 10-NOV-2000; 2000GB-0027549.  
XX 22-DEC-2000; 2000GB-0031706.  
XX (CHIR-) CHIRON SPA.  
XX Ratti G, Grandi G;  
XX WPI; 2002-154726/20.  
XX N-FSDB; ABL91265.  
XX Novel Chlamydia pneumoniae protein useful in the manufacture of a  
medicament for treatment or prevention of infection due to Chlamydia,  
preferably Chlamydia pneumoniae, and for diagnostic purposes -  
XX Claim 1; Page 120-121; 364pp; English.  
XX Sequences ABB90526-ABB90715 represent novel proteins from Chlamydia  
pneumoniae (strain CW029), and ABL91184-ABL91373 represent DNA encoding  
them. The proteins are predicted to be immunogenic and may therefore be  
XX useful in vaccine production and for diagnostic purposes. Chlamydia  
XX pneumoniae is a common cause of respiratory disease in humans, and is  
XX also involved in the development of cardiovascular diseases such as  
XX atherosclerosis, coronary artery disease, carotid artery stenosis,  
XX myocardial infarction, cerebrovascular disease, aortic aneurysm,  
XX claudication and stroke. The proteins and nucleic acids of the invention  
XX may be used in vaccines and pharmaceutical compositions for the  
XX prevention or treatment of chlamydial infections, particularly Chlamydia

CC pneumoniae infections. The proteins may also be used in the detection of  
CC Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched  
CC DNA probe assay or blotting techniques for determining Chlamydia  
CC pneumoniae gene expression. The present sequence represents a  
CC specifically claimed Chlamydia pneumoniae protein of the invention.  
XX  
SQ Sequence 754 AA;

Query Match 5.58; Score 212.5; DB 23; Length 754;  
Best Local Similarity 21.3%; Pred. No. 56-07;  
Matches 111; Conservative 82; Mismatches 217; Indels 111; Gaps 16;

QY 287 QKVTLRNLNDTQIIITAGNWLNVKSAAPGYFTFQVLPKQNLSEGGVNNAPKFTTGR 346  
Db 234 QHVLKKFNPETHVDVIAGRVIFGSAGEVG----ELLKIYFVQSESIROYRVIPLT 289  
QY 347 KISLDFODVEIRILQILAKESGNIVASDSVNGKWTLSLKDPWDO-----ALDL 397  
Db 290 KI-----DPGEMISILNAFREDITKDVSEES-----LGLRVVLPLOYQGRSLFSGTAAL 339  
QY 398 VMOARNLDMRQGVNVIAPREDLLAKDAFLQAEKDIALGALYSQNFOLKYNVEEFR 457  
Db 340 VQQAALTIRELEGIEN--PTDKTVF---WYNVYKSDPQELALLSQ-----VHDVFSGE 389  
QY 458 SILRLNADTTGNR-----NVLVSGR-----GSVLIDPATNLIVTDRSVIEKF 502  
Db 390 NKASVGAADCGSQLNASIQIDITTVSSAKDGSVKYGNFIADSKTCLIMWVEKEVLPRI 449  
QY 503 RKILDELDPAAQVMIEARVEAADGFSRDLGVKFGATGKKLKNBTSAFGWGWSGFGG 562  
Db 450 QMLLKLKLDVKKVRIEVLVFERKLAHEQKSLNLLRLEEVCKGCS-----497  
QY 563 DRKGAETKINLPITAAANGISILVRAISSGALNLELSASELS-----KTKTLANPR 614  
Db 498 -----PSVSWAGGTGILEFLFKGSTGSSIVPGYDLAYQFLMAQEDVRINASPS 545  
QY 615 VLTONRKEAKIESGEYIPTFTVSIANGSGSNTTELKKAIVGLTV--TPNIT---PDGQII 669  
Db 546 VVTWNTQTPARIAVDEM-----SIIVSSDKDKAQYNAQYGINIKMLPVINUGEECKSY 600  
QY 670 MTV-----KINKDSPAQASGNQITLCISTKNLNTQAMVNGGTLIVGGIYEEDN 719  
Db 601 ITLETIDITFTGKNHDDRDP-----VTRRNITNKVRIADGETVIIGLRCQKM 649  
QY 720 GNILTKVPLGDPVIGNLPKTRGKKTDRRELLIFITPRIM 760  
Db 650 SDSHDGPIPLGDIPIGKLFMSSTSDSLTEMPVFITPKIL 690

RESULT 24  
ABB47325  
ID ABB47325 standard; Protein; 1711 AA.  
XX ABB47325;  
AC ABB47325;  
XX  
DT 05-FEB-2002 (first entry)  
XX  
DE Listeria monocytogenes protein #29.  
XX  
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
XX vitamin B12; bacterial infection; disease.  
XX  
OS Listeria monocytogenes.  
XX  
PN WC200177335-A2.  
XX  
PD 18-OCT-2001.  
XX  
PF 11-APR-2001; 2001WO-FR01118.  
XX  
PR 11-APR-2000; 2000PR-0004629.  
XX  
PA (INSP ) INST PASTEUR.

XX Buchrieser C, Frangeul L, Couve E, Rueniok C, Fsihi H, Dehoux P;  
PI Dussurget O, Chetoui F, Nedjari H, Glaeser P, Kunst F, Cossart P;  
PI Daniels J, Goebel W, Kretz J, Kuhn M, Ng E, Vazquez-Boland JA;  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Anand L;  
PI Chakraborty T, Domann B, Hain T, Berche P, Charbit A, Durant L;  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
PI Rose M, Voss H;  
XX  
DR WPI: 2002-010914/01.

Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
PT and prevention of Listeria and related bacterial infections, and  
PT related polypeptides

XX Claim 6; SEQ ID No 30; 192pp; French.  
CC The present invention relates to the genome sequence of Listeria  
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
CC it are useful for selecting probes and primers for detecting genes in L.  
CC monocytogenes and related organisms, and for studying genetic  
CC polymorphisms and other genomes. The present sequence is a protein  
CC encoded by the genome sequence of the present invention. Proteins  
CC expressed from the genome sequence are useful for raising specific  
CC antibodies, identification of L. monocytogenes and related organisms, and  
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
CC B12. The genome sequence and proteins encoded by it are also useful for  
CC selecting compounds that regulate gene expression and cell replication  
CC and modulate L. monocytogenes-related diseases. In addition, the genome  
CC sequence and proteins encoded by it are useful in pharmaceutical and  
CC vaccines compositions for the treatment or prevention of infections by L.  
CC monocytogenes and related organisms.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1711 AA;

Query Match 4.8%; Score 183; DB 23; Length 1711;  
Best Local Similarity 20.3%; Pred. No. 0.00031;  
Matches 176; Conservative 116; Mismatches 315; Indels 258; Gaps 44;

QY 9 ISGLFVATAAQTASAGNITDKVSSLPNKQKI---VKVSFDKIVNPTGVTSSPARIA 65  
Db 959 ITSDPADQVKFTPGVYAT--LKANDISGNQAIPTVNVTI--KVTAPEV---ITADEA--- 1011  
QY 66 LDPEQTGIMDQVLEYADPLLSKISAAQN-----SSRRLVNLNKPQY-----NTEV 115  
Db 1012 -----IIEKGITKTSATDELTDIHTTNDGSAITSNFATVVDLNTPGDYDVLSTDT 1064  
QY 116 RGNKWIWIFINESDDTVSAPAPVAKAAPAKQCGCTVYQVRSIRIQTLYPGKTTAA 175  
Db 1065 GGN-----ISDPVITITVTRDTTPPVITADTTI-----TYEKGTTKTA 1103  
QY 176 PF-----TESVSVSAPFSP--AKQ-----QAAASAKQOT---AAPAKQOTAAPAKQOA 219  
Db 1104 AFLTDVNATTNDGSAVTSNFPISLKQVGTQVTLSSVDENGVALPKVTVVQDTQKP 1163  
QY 220 AAPAKQTNIDFRKQOK---NAGITELALGAGAGPDISQQHDHII-----VTLK-- 265  
Db 1164 VISSATSITYERGEKKQETGFLDLAIKTDDGTP-VTSDFDTAVNLDTAGTYVTLNSE 1222  
QY 266 ---NHTLPTTLQSRSLDVADEKTPV-----QKVTLKEL-----NNDTQLII 302  
Db 1223 DESGNKADPVTI--TVTVADTEKPIITADTTITVAKGTTKTVAQFLTDIHTTNDGSTII 1280  
QY 303 TT-----AGNWELV-----NKSAPGYFTFQVLPKQNLSEGGVNNAPKFTT--- 344  
Db 1281 SNFDPVILAQEGTYTVVLNAKDESNEADPVTITVVDTK-----GPIINALNAITVER 1335  
QY 345 -----GRKISLDFOVDIRILQILAKESGNIVASDSVNGKWTLS 385

Db 1336 TINKHEADFLADIEATDDGSTITTDFNSKDLDTV-----GTIV-----TLN 1378  
 Qy 386 LKQVWQALDLVQARNLDMRQGNVNIAPRDELLAKDKAFLOAEKDI-ADLGALYSQ 444  
 Db 1379 AEDASGNKATPV-----KVIKVEDTIPPIITADQSITYERGITKTEQAFTYDIKAATSD 1433  
 Qy 445 NFQLYKNVVEFRGILRLDNADTTGNENTLV-----SGRGSVLIDPATNLIIVTDRSVI 499  
 Db 1434 NSPIS-----SDFSKI-----DLTKGNYEVLLRATDQSGNKAL---PLKINVLQDITIAPV 1482  
 Qy 500 EK--FRKILDELDPVQAQVMIEARIVAAADQFSRDLGVKGATGKKLKN---DTSAGFW 554  
 Db 1483 IKTTTSREITABRGTPMTEQQLLAKI-----GANTDDGSKITTDVNPAINVNTSGDYLVLHY 1537  
 Qy 555 GWSGFGDDKMGAEETKINL-----PITAAANSIS-----LVRASSGALNLEISASSELS 605  
 Db 1538 AVDA--AGNQAIPEVETIHKVDITAPIKADKKSYPGVTKVTETFLQDIHATDDG-S 1594  
 Qy 606 KTKTLANPRVLTQNRKEAKIESGVEIIPFTVTSIANGSGSSTNTELLKAVLGLTVPNTIP- 664  
 Db 1595 KITTFDFENML-----KTPGKYTIHLNAVDADGNKAKTIDVSLTVEEKVTPP 1641  
 Qy 665 -----DQGIIMTVKINKDSPAQCASGNQITLIC:STKNLNTQAMVENGSGTLIVGGIYEED 718  
 Db 1642 KPPTDDGG-----NNTNGSNGTNGSNTDITVNPTKQNTATNES----- 1681  
 Qy 719 NGNTLTQVPLLD-----IPVTGNLF 739  
 Db 1682 -----IPALGDTKSTIEPVGME 1699

RESULT 25  
 ABG18141  
 ID ABG18141 standard; Protein; 671 AA.  
 AC  
 AC ABG18141;  
 DT 18-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #18132.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 XX WO200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AA982328.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX Claim 20; SEQ ID No 48500; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (II) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags  
 for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 671 AA;  
 Query Match 4.7%; Score 182; DB 22; Length 671;  
 Best Local Similarity 23.1%; Pred. No. 9.4e-05;  
 Matches 82; Conservative 53; Mismatches 110; Indels 110; Gaps 13;  
 Qy 168 PG-KTTAAAPFTESVVSAPFSPAQQAAAGAKQQTAAAPAKQQAAPAKOT 226  
 Db 405 PGVPSQSAGPGSGSVVDV-----PVVPMVSGKTPALSAQATALMT 446  
 Qy 227 NIDFRKDGKNAGIIEALGAFAGQDPDISOOHDIIVTLKNHTLPTTLQSLDVADPKTPV 286  
 Db 447 YLDER-----PDVS-----SLDVG-YSLAL 465  
 Qy 287 QKVTAKR---LNNDTQ-----LIITAGNMLVNKSAAPGVFTFOVLPKKNLESQGV 336  
 Db 466 TRSALDERAVLGSRETLICDVLITGTNCE-----PPQWLWQRQGDTEAMDRTT 519  
 Qy 337 NNAPKTTFTGRKISLDFQDVEIRTILOILAKESGMNIVASDSVNGKMTLSLKDVWDQALD 396  
 Db 520 VDADTERTGTGKAASDADAQ---VLOALAEQEKLNIVSPVSGTIVSLHITDVPWKQALQ 576  
 Qy 397 LVWQARNLDMRQGNIV-----NIAPRDELLAKDKAFLOAEKDIADLGALYSQNF 446  
 Db 577 TVVKSAGLITRQGNILSVHSIAWQNNIARQEAQRAQANLPE-----NRSI 626  
 Qy 447 QLYKNVVEFRSILRLDNADTTGNRNTLVSGRGSVLIDPATNLIIVTDRSVIEK 501  
 Db 627 TLQYADAGELAK-----AGEK--LLSAKGMTVDRKTRNLLRLDRDNKTVGR 670

RESULT 26  
 ABJ19119  
 ID ABJ19119 standard; Protein; 10498 AA.  
 XX  
 AC ABJ19119;  
 DT 06-MAR-2003 (first entry)  
 XX Pathogen specific antigen related staphylococcal protein SEQ ID No 440.  
 DE Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;  
 XX hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;  
 KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;  
 KW autoimmune disease; HIV; hepatitis.  
 XX Staphylococcus sp.  
 XX WO200259148-A2.  
 XX 01-AUG-2002.  
 PD 21-JAN-2002; 2002WO-EP00546.  
 PF 26-JAN-2001; 2001AT-0000130.  
 PR











pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pat\_sequences.

Query Match 4.3%; Score 166.5; DB 22; Length 5795;  
Best Local Similarity 19.9%; Pred. No. 0.032;  
Matches 175; Conservative 117; Mismatches 324; Indels 265; Gaps 37;

CC 474 LVSGRGVULDPATNLTIVTDRSVEIKF-----RKLDELDPVPAQVMIE 519  
CC 1004 LOSSVNVQV---ESTAGM---TQOSIDNNAKKREAEITEITAAQRVINGDRTAQQISDE 1056  
CC 520 ARIVEAADGFSRDLGVKFGATGKGLKNDTSFAGWGVNSFGGDDKWAET-KINLPITA 578  
CC 1057 KHRVDNA-----LTALNQAKHDLTADTHALEQAVQQLNRTGTTGCKPASITA 1104  
CC 579 AANSISLVRAISSGALNLSSESLSK-----TKTLANPRVL 616  
CC 1105 YNNSI---RALQSDLTSAKNSANAIQKPIRTVQEVQSALTNNRVNRELTAQINOLVPL 1161  
CC 617 TQNR--KEAKIESGYEIPFTVT-----SIANGSSSTNT 647  
CC 1162 ADSALKTKATKLDSEINKSVITDGMTQSSIQAYENAKRAGQTESNAQVNNNGDATDQ 1221  
CC 648 E-----LKKAVLGLVTTPNTPDQGIIMTVKI-----NKDSP-----AQ 681  
CC 1222 QIAAEKTKVEEKNSLQKQIAGL--TPDLAP-----LQTAQTQLQNDIDQFTTTGWTSSAS 1275  
CC 682 CASSNQITLILCISTK-----NLNTQAMVENGGLIIVGGIYEEDNGNTLTQVPL 728  
CC 1276 IAAFNEKLSAARTKIOEIDRVVLASHEDVATIRQNVTAANAASALDQARNGLTVDKAPL 1334

RESULT 31  
AAU37017  
ID AAU37017 standard; Protein; 5795 AA.  
XX AC AAU37017;  
XX DT 14-FEB-2002 (first entry)  
XX DE Staphylococcus aureus cellular proliferation protein #1187.  
XX KW Antisense; prokaryotic cellular proliferation protein;  
XX KW antibiotic; antibacterial; drug design.  
XX OS Staphylococcus aureus.  
XX PN WO200170955-A2.  
XX PD 27-SEP-2001.  
XX PF 21-MAR-2001; 2001WO-US09180.  
XX PR 21-MAR-2000; 2000US-191078P.  
XX PR 23-MAY-2000; 2000US-206848P.  
XX PR 26-MAY-2000; 2000US-207727P.  
XX PR 23-OCT-2000; 2000US-242578P.  
XX PR 27-NOV-2000; 2000US-253625P.  
XX PR 22-DEC-2000; 2000US-257931P.  
XX PR 16-FEB-2001; 2001US-269308P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
XX PI Yamamoto RT, Xu HH;  
XX DR WPI; 2001-611495/70.  
XX DR N-PSDB; AAS54876.  
XX PT New polynucleotides for the identification and development of  
XX PT antibiotics, comprise sequences of antisense nucleic acids -  
XX PS Example 3; Seq ID No 12610; 511pp; English.  
XX CC The invention relates to antisense inhibitors of genes essential to  
XX CC prokaryotic cellular proliferation, their use in identifying the  
XX CC genes, their use in the discovery of novel antibiotics, the essential  
XX CC genes themselves and the encoded proteins. The prokaryotes used are  
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pat\_sequences.

Query Match 4.3%; Score 166.5; DB 22; Length 5795;  
Best Local Similarity 19.9%; Pred. No. 0.032;  
Matches 175; Conservative 117; Mismatches 324; Indels 265; Gaps 37;

QY 34 SLPNKQKIV-----KVSFKEIVNPTGFTVTSPPARIALDFEQTGISMDOOVLEYADPL 86  
DB 2387 TIANKPDVTVLDAHTGKVTFNANTVKPNSAITITPK-----AGTG----- 2426  
QY 87 LSKISAAQNSSSEARLVNLNKPQVNTVEVRGNKVMIFINESDDTVSAPARPAVKAAPAA 146  
DB 2427 ---HSASNP-----TLTAPATHV-----NTTEIVKDYGSNVTAAEINNVAQVA--- 2469  
QY 147 AKQGCRTVYQVRSIRIQTLYPGKTTAAAPT-----ESVSVSAPSPAKQQ 194  
DB 2470 ---NKRTATIKNGTAMPTNLGGSTTIPVTIVYNDGSTBEVQESI-----FTKADKR 2519  
QY 195 AAASAKQQTAA-----AKQO---TAAPAKQO-----AAPAKOTN 227  
DB 2520 ELITAKNHLDDPVSTGKKPGTITQYNNAIHNAQOQINTAKTEAQVNNERATPOQVSD 2579  
QY 228 IDFRKDGKNAGIIELAALGFAGQPDISQCHDHIIVTLKN-----HTLPTT---LQRLSD 278  
DB 2580 ALTKVRAAQTKINEAKAL-LQNKEDNSQ-----LVTSKNLQSSVNVPSITGMTQCSID 2633  
QY 279 VADFK-----TPVQKV-----TLKRLNNDTQLIITTAGNWLNVK-SAAPGYFTFO 323  
DB 2634 NYNAKKREAEITEITAAQRVINGDRTAQQISDEN-----TTAQRNDLTNQISQATNLAVE 2689  
QY 324 VLPKQNLSEGGVNN---APKFTGKISLDPQDVE-----IRTLQILAKESGM 370  
DB 2690 SVKQSANSLDGAQNLQTAINDKSGTSLASQNFDADEQKNAYNAQVNAETILNKQTGP 2749  
QY 371 NIVASDSVNGKMTLSLKVDPWDQALDLVNAQ-----NLMROQGHIVNTAPRDELLA 423  
DB 2750 N-TAKTAV-----EQALNNVNSAKHALNGTQNLNNAKQAAITAINGASDLNQ 2795  
QY 424 KQKAFLOAEKDIALGALYSQNFOLKYKIVEFRSIL-----RLDNAD 466  
DB 2796 KORDALKAOAN-----GAQVSNADQVORNATELNTAMGTCLKHAIADKNTNLASSKYVNA 2851  
QY 467 TT-----GNRNTLVSGRGSVLIDPATNLTIVTDRSVEIKF----- 503  
DB 2852 STKQNAVTTKVTNAEHIISGTPVTVPSEVTAANQVNSAKQELNGDERLRAKQNA 2911  
QY 504 --KLIDELDPVPAQVMIEARIVEA-----ADGFSRDLGVKFG 538  
DB 2912 AIDALTQLANTP-QKAKLKEQVQGNRLEDIQVTQNGQALNNAMKGLRDSIANETVKAS 2970  
QY 539 ATGKKKLKNDTSFAGWGVNSFGGDDKWAETKINLPITAAANSISLVRAISSGALNLEL 598  
DB 2971 QNYTDSPPNQSTYNSAVNSAKNGIINQTNPTMDTSAITQATQVNNAKNGLNGAENLN 3030  
QY 599 SASESLSKTKTLANPRVLQNKREA---KIESGYEIPFTVTSIANGSGSTNTLTKAVLG 655  
DB 3031 AQNTAKQNLNTLISH---LTNNOKSAISSOIDRAGHVS-EVTAAKNAATELNTQMNLEQA 3086

|    |      |   |      |
|----|------|---|------|
| Qy | 656  | LTVPIHITPGQILMTVKINKDSPAQCASGNQILCISTKNLT-----QAQVENGSTLI | 7110 |
|    |      | :           |      |
| Db | 3087 | IHDQNTVKQGVTFDADKARDAATNAVSRAETIL-NKTGANTSKODVEAIQN-----  | 3140 |
| Qy | 711  | VGGIYEEDNGNTLTKVPLGDIPIVGNLFMTRGCKTDRREL                  | 751  |
| Db | 3141 | -----VTSAKNALNGDNV-----TNAGNTAKHAAL                       | 3165 |

RESULT 32

ABG11810  
ID ABG11810 standard; Protein; 5464 AA.

AC ABG11810;

18-FEB-2002 (first entry)

Novel human diagnostic protein #11801.

XX Human; chromosome mapping; gene mapping; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens

XX  
PN  
W0200175067-A2XX  
11-OCT-2001XX  
30-MAR-2001: 2001WO-US08631. PEXX  
PR 31-MAR-2000: 2000US-0540217.

PR 31-MAR-2000; 2000US-0340217.  
PR 23-AUG-2000; 2000US-0649167.

XX  
PA (HYSE-) HYSEQ INC.XX  
PI  
Drmanac RT, Liu C, Tang YT;

AA  
DR WPI: 2001-639362/73.

DR WFI; ZOOI-635362/  
DR N-PSDB: AAS75997.

xx New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published/pct](http://ftp.wipo.int/pub/published/pct) sequences.

|    |                   |
|----|-------------------|
| AA |                   |
| SQ | Sequence 5464 AA; |

Query Match 4.3%; Score 165; DB 22; Length 5464;  
Best Local Similarity 19.7%; Pred. No. 0.038;

| Matches | 152; | Conservative  | 115;  | Mismatches             | 296; | Indels | 208; | Gaps | 36 |
|---------|------|---|---|------------------------|------|--------|------|------|----|
| QY      | 62   | ARIALDFEQTGISM---   | DOOVLEYADPL-----                              | LSKISA-IQNSRRARLVNLNKP | 109  |        |      |      |    |
| Db      | 353  | SKIRLNDAAQAKAMARDNKNLDYLEOSGTHKARDLEKMAQSGNQOQLLEITKALTKPR    | 3612  |                        |      |        |      |      |    |
| QY      | 110  | Q-----  | YNTVEGRGNKVM-I-FINESDDTVSAPARPAKAAAPAKQOQGRTV | 155                    |      |        |      |      |    |
| Db      | 3613 | KEGELPNLSAALGYNALTNGEDTGIOVSERD--IAAEANPAYSGLOSS-----         | 3659  |                        |      |        |      |      |    |
| QY      | 156  | YQVRSIRIQTLYPEGKTTAAAPTSESVWSAPFSPAKQQAASAKQOOTAAPAKQOOTAAPA  | 215   |                        |      |        |      |      |    |
| Db      | 3660 | -QFDPTRDPALNPGIRLDCARYVQSECD---ELLEDRQDALAMAAQAGCLKRYLSITL    | 3715  |                        |      |        |      |      |    |
| QY      | 216  | KQQAAPAKATNIDFRKDGNGAGITELAAAGFAGQPDISQQOHHIIVTLKNHTLPTTLQR   | 275   |                        |      |        |      |      |    |
| Db      | 3716 | QKAEATLVGVSDIENALDVAROTPMTEAAKTAS--DIFKMSDEEILNMPG---PSSIQN   | 3770  |                        |      |        |      |      |    |
| QY      | 276  | SLDVADFKTPVQKVTKLRLNNDTQLIITAGNWLNVKSA-----                   | POYFTFOVLPK                                   | 327                    |      |        |      |      |    |
| Db      | 3771 | NPEVNNNTPAQDPPIVEVEQTEDEPVTTEPT-EVENTSESINSDAFTTGNDSATVDPN    | 3829  |                        |      |        |      |      |    |
| QY      | 328  | KQNLES-GGVNNAKPTFTGRKISLDFQDVEIITLQILAKESGMNIVASDSVNGWNTLSL   | 386   |                        |      |        |      |      |    |
| Db      | 3830 | KVDSNSKPAIDNAPST-----   | TEQQAQVVEDKQVSDOKT---                         | 3864                   |      |        |      |      |    |
| QY      | 387  | KDVPWQALDVLVQARNLDMRQOGNTVNIAPRDELLAKDKAPLQAEKDIADLGALYSQNF   | 446   |                        |      |        |      |      |    |
| Db      | 3865 | -DFDYKAGYERIM---APFKANGKMITPOSPEEVIS-----                     | LMQWGA---                                     | NY                     | 3904 |        |      |      |    |
| QY      | 447  | QLYKNVVEERSILRLDNDATTGNRNTLVSGSGSVLIDPATNTLVITDTRSVIEKFKL-    | 505   |                        |      |        |      |      |    |
| Db      | 3905 | TRKQEIOPVKKYLLM-----  | LENNGLDEBQGLSVAID-----                        | LVGNPAIDAKKLKVES       | 3953 |        |      |      |    |
| QY      | 506  | -IDELDV-PAQQYMIEA---RIVEAADGFSRDL-GVKFGATGKKLK-----           | NDTSAPG                                       | 553                    |      |        |      |      |    |
| Db      | 3954 | GIDVLIDINPEEVKYQAGNHRVSDAAAFATELEDAKSTPEQATLQIISTITWDDHKSQA   | 4013  |                        |      |        |      |      |    |
| QY      | 554  | GWVNSGFGDDDKWGAETKINLPITAAANSI-----                           | SLVRAISSGALNLE-----                           | 597                    |      |        |      |      |    |
| Db      | 4014 | LLENRGL-----  | VNTIVEQKANGIYDRIVEIERLKIQCIIPIGTAFDAYNQV      | 4062                   |      |        |      |      |    |
| QY      | 598  | ---LSAS---ESLSKTXLANRVLTLQ-----                               | NRKEAKIESGYEIPFTVT-                           | 636                    |      |        |      |      |    |
| Db      | 4063 | GNVLAANGGFNDLIKQKPVQEPVVPVQVATRVATPKSEIITNTQAAAAA-----        | PSRVTP  | 4118                   |      |        |      |      |    |
| QY      | 637  | -----STANGSSTNTE-----   | LKAVLGLTVTFNITPDQII-----                      | MTVKINKOSPA            | 680  |        |      |      |    |
| Db      | 4119 | RKTEKSIDGANSQDMQTFWFLKKAIIITARKSEQYFNPPLASVTNMPKHVYKTKIVYVFP- | 4177  |                        |      |        |      |      |    |
| QY      | 681  | QCASGNQTLICISKINLTQAMVENGOTLIVGQIY--EEDNGNTLTKVPLL            | 729   |                        |      |        |      |      |    |
| Db      | 4178 | -----LLDPRNINDOGIDASGATVNGNLVGSSKDIGNITSKLPLL                 | 4218  |                        |      |        |      |      |    |

RESULT 33

ABGI1811

ID ABGI1811 standard; Protein; 7718 AA.

RESULT 33

RESULT 33  
ARG17817

ABG11811  
ID ABG11811 standard; Protein; 7718 AA.

AA ABG11811;

18-FEB-2002 (first entry)

Novel human diagnostic protein #11802.

AA Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder;  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
AA food supplement; medical imaging; diagnostic; genetic disorder;

XX  
OS  
Homo sapiens.

XX PN WO200175067-A2

XX  
PD  
11-OCT-2001

XX 30-MAR-2001; 2001WG-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS75998.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX Claim 20; SEQ ID No 42170; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcr\_sequences.  
XX Sequence 7718 AA;  
SQ  
Query Match 4.3%; Score 165; DB 22; Length 7718;  
Best Local Similarity 19.7%; Pred. No. 0.064;  
Matches 152; Conservative 115; Mismatches 296; Indels 208; Gaps 36;  
QY 62 ARIALDFEQTGISM---DOQVLEADPL-----LSKISA-AQNSRARLVNLNKP 109  
DB 1232 SKIRLNDAAQAKAERDNKNLDYEGESGTHARDLEKMAQSQGNOQLEITKALTKPR 1291  
QY 110 Q-----YNTVEVRGNKVM-I-FINESDOTVSAAPAPAVKAAPAPAKQOQCRTV 155  
DB 1292 KEGELPPNLSAAGYNALTNGEDTGIQSVSRD--IAEANPAYSLGSS----- 1338  
QY 156 YQVRSIRIOTLYPGKTTAAAFTESSVSVSPFSPAKQQAASAKQOATAAPAKQOATAAPA 215  
DB 1339 -QFOPTROPALNPGIRLDCARYQVQESCD---PLLEDQRADALAMAAAGCLKEYLEITL 1394  
QY 216 KQRAAPAKQTNIDFRKDGKNAGIELAALGFAGQPDISQCHDHIIVTLKNHTLPTTLQR 275  
DB 1395 QAAETLVGVQSDIENALDVARDTPTMTFAKTAS--DIFKMSDEILNMPG---PSSIQN 1449  
QY 276 SLVDADFPTPVQVKTLKRLNNDTQLIITAGNWLNVKSA-----PGYFTFQVLKP 327  
DB 1450 NPEPVENNTPAQPDPIVEQTEDEFEVTEPT-EVENTSESESNDAFTTNGDSATVDPN 1508  
QY 328 KQNLFS-GGVNNAKPTFTGRKISLDFQDVE-RTILQILAKESGNIVASDSVNGKMTLSL 386  
DB 1509 KVDNSKSKPATDNAPT-----TEGQAKVVEDKQVSKQKT--- 1543  
QY 387 KDVPWDQALDLVWQARNLDMRQGNINVIAPRDELLAKDKFLAQEKDIADLGALYSQNF 446

DB 1544 -DFDYKAGYERIM-----APFKANGKMITPQSPPEVIS-----LMOMGA---NY 1583  
QY 447 QLKYNVVEEFRSILRLDNADTTGNRNTLVSGRGSVLIDPATNTLIVTDTRSVIEKFKL- 505  
DB 1584 TRKQEIQPYRKYLMM-----LENNGLLDEQQLSVAID-----LVKGNPDAIKKLVKES 1632  
QY 506 -IDELDV-PAQVVMIEA---RIVEAADGFSRDL-GVKFGATGKKKIK-----NDTSAFG 553  
DB 1633 GIDVLDINPEEVKYQAGNHRVSDAEAAFAFELEDAKSTPEQATLQLISTWDDHSKQA 1692  
QY 554 WGVNSGFGGDDKWAETKINLPITAAANSI-----SLVRAISSGALNLE----- 597  
DB 1693 LLENRGL-----VNTIVEKANGIYDRIVSEIERLKILQIPGTAFLDAYNOV 1741  
QY 598 ---LSAS---ESLSKTKTLANPRVLTQ-----NRKEAKIESGYEIPFTVT- 636  
DB 1742 GNYLAANGGFNDLIKQKEVQPEPVQVATRPATPKSEITNTQAAAAA-----PSRVTP 1797  
QY 637 -----SIANGGSSTNTE-----LKKAVLGLTVPNTITPDQII-----MTVKINKDSPA 680  
DB 1798 RTEKSSIDGANSDDQMTFFWLKKAIITARKEQYFMPPLASVTNMPKHYGKTIKVEYVP- 1856  
QY 681 QCASGNQITLCISTKVLNTQAMVENGGLIVGGIY--BEDNGNLTQVPLL 729  
DB 1857 -----LLDDRNINDQGDASGATVANGNLYGSSKDIGNITSKLPLL 1897  
RESULT 34  
AAB01837  
ID AAB01837 standard; Protein; 1073 AA.  
XX AAB01837;  
XX AC AAB01837;  
DT 11-SEP-2000 (first entry)  
XX Haemophilus influenzae strain LCD2 mature HMWZA protein, SEQ ID NO:49.  
DE Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;  
KW non-typeable Haemophilus influenzae; NTH1; non-encapsulated;  
KW recombinant production; Escherichia coli; antibacterial; vaccine;  
KW human disease; Otitis media; epiglottitis; pneumonia; tracheobronchitis;  
KW detection; diagnosis.  
XX Haemophilus influenzae strain LCD2.  
OS WO200020609-A2.  
XX PN 13-APR-2000.  
PD 07-OCT-1999; 99WO-CA00938.  
PF 07-OCT-1998; 98US-0167568.  
PR 08-DEC-1998; 98US-0206942.  
XX (CONN-) CONNAUGHT LAB LTD.  
XX Loosmore SM, Yang Y, Klein MH;  
PI WPI; 2000-303789/26.  
DR N-PSDB; AAS52186.  
XX Nucleic acid molecule for producing recombinant high molecular weight  
PT proteins of Haemophilus which are used as a vaccine to provide  
PT protection against Haemophilus induced diseases in humans -  
XX Claim 8; Fig 23A-E; 307pp; English.  
XX The invention relates to the recombinant production of Haemophilus  
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The  
CC expression construct used to effect recombinant expression comprises a  
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked  
CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.

```
CC influenzae (NTHi). Most hmw-expressing NTHi strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural hmwA proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the hmwA
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature hmwA. The invention also discloses hmwA genes (AAAS2175-AS2198)
CC and hmwA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC strains Joyce, K1, K21, LDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae hmw
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The hmw proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC hmw proteins and/or hmw peptides. The nucleotide sequences encoding the
CC hmw proteins can be used to isolate and clone hmw genes from other
CC non-typeable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents a mature hmwA protein from a non-typeable
CC strain of H. influenzae.
XX
XX Sequence 1073 AA;
Query Match 4.2%; Score 160.5; DB 21; Length 1073;
Best Local Similarity 19.7%; Pred. No. 0.0075;
Matches 125; Conservative 88; Mismatches 212; Indels 211; Gaps 25;
QY 205 APAKQTAAPAKQAAPAKQTNIDFRKDGKNAGIIEALALGFAGQPDISQDHHIIVTL 264
DB 8 APSAERT-----DTGEDVEYTGADII-----NHQKQNETSKTL 42
QY 265 KNHLPPTLQSLVDADKPTVQKTLKRLNNDLQIITAGNWLKNSAAPGVFFQV 324
DB 43 TTTLEGLMKGL-----FVNITARKIRVNSTINIGDSGHLT 80
QY 325 LPKKNLSEGGVN-NAPKTFTEKISLDFQD-VEIRTILOLAKESGMNIVASDV----- 378
DB 81 LYKKRNRSDG:QINKDITSTGSLTINSDDWDVHG--NITLGBGLNITSDSVAPEG 138
QY 379 -----NGKWTLS-----LKDVPWD---QALDLVMOARLNRQQNI 412
DB 139 GNGNKRSSASAQI:TAQGTTITGKNTFRLNKNSVLTNGTGLSITSTASNLHRDGEI 198
QY 413 VNIAPR---DELLAKDKAPLOAKEDIDALGYSONFOLKYKNVEEPRSIILDLNADT-- 467
DB 199 -NVSGNVTINQTTQNIQVWKASSD-----SYNVTFSN--LREDSKFTFI 241
QY 468 ----TGURNTLVSRGSLVDIPATNLTIVDTRSVIEKFKRLIDELDVQAQWIEARIVE 524
DB 242 KYVNSARNGDVRGRSF-----AGVIF 262
QY 525 AADGFSRLDGVKFGATGKKLKNDTSAFGWVNSGFGGDKWGAETKINLPITAAAN-SI 583
DB 263 NAKGLTTSFNVKGSIVDFKLK-----FNSGVNSQKRIPIQFSNISVSGGRVNI 313
QY 584 SLVRAISGALNLELSASSLSKTKTL-----ANPVLITQ- 618
DB 314 NTLANLTGGVEIR-SSSINVSDGSLTMTAQARDNAFEITKDLVINASNLNLSIIQON 372
QY 619 ----NRKEAKTESGVEIPFTVTSIANGSGSTNTTELKAVLGLTVPNTPDGQIMTVK 673
DB 373 DGFDDNQKANA NSKNVITQGGNVTLGGQSSSITGVS-----NIGANANVTLOAH 425
QY 674 INKDSPAQASGNQTI-----LCISTKNLTNQAMVENG-----TLIVGGIYEEDNG 720
DB 426 NGNDRNKKLTFFGNVSVEGELRLVGASANNLNLSVKSAGAKFAETNDNLNITGTF-TNNG 484
QY 721 NTLTKV-----PLIGDIPVIGNLFKTRGKKTDRRELL 752
DB 485 TSIIDVKKGAUKLGNITNDGNLNIITNAKNGQKSVI 520
```

```
RESULT 35
AAB01836
ID AAB01836 standard; Protein; 1079 AA.
AC AAB01836;
XX 11-SEP-2000 (first entry)
DE Haemophilus influenzae strain LDC2 hmw2A protein, SEQ ID NO:47.
XX
XX hmw protein; hmw gene; hmwA1; hmwA2; high molecular weight;
XX non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
XX recombinant production; Escherichia coli; antibacterial; vaccine;
XX human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
XX detection; diagnosis.
XX
XX Haemophilus influenzae strain LDC2.
XX
XX WC2000020609-A2.
XX
XX 13-APR-2000.
XX
XX 07-OCT-1999; 99WO-CA00938.
XX
XX 07-OCT-1998; 98US-0167568.
XX
XX 08-DEC-1998; 98US-0206942.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Loosmore SM, Yang Y, Klein MH;
XX
XX WPI: 2000-303789/26.
XX
XX N-PSDB; AAAS2185.
XX
XX Nucleic acid molecule for producing recombinant high molecular weight
XX proteins of Haemophilus which are used as a vaccine to provide
XX protection against Haemophilus induced diseases in humans -
XX
XX Claim 12; Fig 23A-P; 307pp; English.
XX
XX The invention relates to the recombinant production of Haemophilus
XX influenzae high molecular weight (HMW) proteins in Escherichia coli. The
XX expression construct used to effect recombinant expression comprises a
XX promoter functional in E. coli (e.g., the T7 promoter) operably linked
XX to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
XX influenzae (NTHi). Most hmw-expressing NTHi strains contain two hmw gene
XX clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
XX hmwB and hmwC genes. The hmwA genes encode the structural hmwA proteins
XX and the hmwB and hmwC genes encode accessory proteins which are
XX responsible for post-translational processing and secretion of the hmwA
XX proteins. The modified hmwABC operon used in the expression construct of
XX the invention contains an A gene modified such that it encodes only the
XX mature hmwA. The invention also discloses hmwA genes (AAAS2175-AS2198)
XX and hmwA proteins (AAB01824-B01849) from the non-typeable H. influenzae
XX strains Joyce, K1, K21, LDC2, PMH1, 15 and 12. The nucleic acids and
XX vectors are used for the production of recombinant H. influenzae hmw
XX proteins which can be used as vaccines to mediate a humoral or
XX cell-mediated immune response to provide protection against diseases in
XX humans caused by H. influenzae (e.g., otitis media, epiglottitis,
XX pneumonia and tracheobronchitis). The hmw proteins are also useful as
XX antigens in immunoassays for detecting antibodies against Haemophilus,
XX hmw proteins and/or hmw peptides. The nucleotide sequences encoding the
XX hmw proteins can be used to isolate and clone hmw genes from other
XX non-typeable strains of Haemophilus via hybridisation reactions. The
XX present sequence represents an hmwA protein from a non-typeable strain of
XX H. influenzae.
XX
XX Sequence 1079 AA;
Query Match 4.2%; Score 160.5; DB 21; Length 1079;
Best Local Similarity 19.7%; Pred. No. 0.0076;
Matches 125; Conservative 88; Mismatches 212; Indels 211; Gaps 25;
```



Db 2925 RGSIAADATKQNTVDASQNK---KDYNNAVTT-----AQGIIDOTTSTPT----- 2969  
 Qy 518 IEARIIVEAADGSRDLGVKFGATGKKKUK--NDTSAPGKGVNSGGGDKWGAETKINLP 575  
 Db 2970 LDPTVINOAG--QVSTTTNAGNGNENLEAAKQQAQSLGSLDNLNNAKQTVTDQIN-- 3025  
 Qy 576 ITAAANSISIVRAISSGALNLELS-----ASELSKTKTLANPRVLTONRKEAKIESGY 629  
 Db 3026 ---GAHTVDEANCIKQNAQNLANMGNLQKADKATKATVNFDDAQKQA-----Y 3077  
 Qy 630 EIPFT---VTSIANGSGSTNTLKVAVLGLTVTNITPDGQILMTVTKINDSPAQASG 685  
 Db 3078 NTAVTNAENIISKANGNATQAEVEQAIKQVNAKQ-----ALNGNANVQHAQ 3126  
 Qy 686 NQILICISTKNLN-----TOAMVNGGTLIVGGI-----YEBDNGMTLTQVPLLG-- 730  
 Db 3127 EATALINSSNDLNOAKDALKQOVQNAIT--VAGVNVVVKQTAQELNNATQKQGIADKE 3184  
 Qy 731 DIPVIGNLFTKRGKTD-----RRELLIFITPRIMGT 762  
 Db 3185 QTKADGNFVNADPKONAYNOAVAKAELISATPDVVVT 3223

## RESULT 37

AAU34387

ID AAU34387 standard; Protein; 837 AA.

XX AAU34387;

AC AAU34387;

XX AAU34387;

DT 14-FEB-2002 (first entry)

DE Staphylococcus aureus cellular proliferation protein #663.

XX Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

XX Staphylococcus aureus.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

XX 23-MAY-2000; 2000US-206848P.

XX 26-MAY-2000; 2000US-207727P.

XX 23-OCT-2000; 2000US-242578P.

XX 27-NOV-2000; 2000US-253625P.

XX 22-DEC-2000; 2000US-257931P.

XX 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

XX N-PSDB; AAS52246.

XX New polynucleotides for the identification and development of

XX antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 5883; 51pp; English.

XX The invention relates to antisense inhibitors of genes essential to

XX prokaryotic cellular proliferation, their use in identifying the

XX genes, their use in the discovery of novel antibiotics, the essential

XX genes themselves and the encoded proteins. The prokaryotes used are

XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

XX invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 837 AA;

Query Match 4.1%; Score 159; DB 22; Length 837;

Best Local Similarity 19.1%; Pred. No. 0.0067;

Matches 175; Conservative 129; Mismatches 303; Indels 308; Gaps 42;

Qy 1 MNTKLTIIISGLFVATAAFOTASAGNITDIKVSSLPNKQIV--KVSFDPKIVNPTGFVT 58

Db 25 LNTAMTALKRA--IADKA-ETKASGY---VNADANKRQAYDEKVTAAENIVSGTPTPT 77

Qy 59 SSPARIALDFEQTGISMDQOVLEVADPLLSKISAAQNSSRARLVNLNPKQYNTVEVRGN 118

Db 78 LTPADVT-----NAATQVTNAKTOLN-----GNHNLEVAQK 108

Qy 119 KWIWIFINESDD---TVSAPARPAVKAAPAPAKQOG-CETVVOYRSIR-----IQTLYPG 169

Db 109 NA---NTAIDGLTSLNGPQAKLK-----EQVQATTLFPVOTVRDNTAQTUNTAKG 157

Qy 170 KTTAAAPFTESVVSAPSPAKQ-----QAAASAKQ---QTAAPA-KQOTAPAKQ 218

Db 158 LRDSJA--NEATIKAGQNYTDASQNKQTDYNSAVTAAKAIIGTTPSPMNAQEQINQAKDQ 215

Qy 219 AAPAKQTNIDFRKQKQAGIIEALGFGAPDQISQOHDHIIIVTLKHNHTLPTTLQRLSD 278

Db 216 VTAKQOALN-----GOENL-----RTACTNAKQHLN 241

Qy 279 VADFKTPVQKVTIKR-----LNNDTQLIITAGNWLNVNKSAAAGYFTFQVLPKKONL 331

Db 242 GLSDLTDAQKDAVKRQIEGATHVNEVQ---AQNADALNTAMT-----NL 284

Qy 332 ESG-----GVNNAPKFTGRIKSLDFQVE-----IRTILOILAKESGNNI 372

Db 285 XNGIQDQNTIKQGVN-----FTDADBAKNAYTNAVQAEQILNKAQGN- 329

Qy 373 VASDSYNGKVTLSLKDVPWDQALDLVMOARNLDMRQQGNIVNIAPRDELLAKDKAFLOAE 432

Db 330 TSKDGV-----ETALENVQRAKN-ELNGNQNVAN-----AKTTAK 363

Qy 433 KDIADLGALYSQNFQLYKYNVEFRSLRLDNDADTTGNR-NTLVSGRGSVLIDPATNTLI 491

Db 364 NALNLTISINNAQKEALKSQIEGATTVAGNVQVSTTASELNTAMNLOINGINDEAATKAA 423

Qy 492 VTDTRSV-----IEKFRKLIDELDPVPAQQVMIERARIVEAADGFSRL----- 533

Db 424 LNGTONLEKAKOHANTAI DGLSHLTNAQKEALKQDVQOSTTVAEQAQNEKANNVDAAMD 483

Qy 534 ---GVKFGATGKKK-----LKNDTSAFGWGVNSGFGDDKKGWGAETKINLPITAAANSI 593

Db 484 KLRQSIADNATTKQNYTDASQNKDAYNNAVTTAQGIIDOTTSTPTLDPVINOACQV 543

Qy 584 SLVRAISSGALNLELS---ASELSKTKTLANPRVLT-----QNRK- 621

Db 544 STTKNALNGNENLEAAKQQAQSLGSLDNLNNAKQTVTDQINGAHTVDEANQIKQNAQN 603

Qy 622 -----EAKIESGYEIPFT-----VTSIANGSGSTNTL 649

Db 604 LNTAMNQLKQAIADKADKATVNFDDAQK--QAYNTAVTNAENIISKANGNATQAEV 662

Qy 650 KKAVLGLTVPNTPDGQIIMTVKINKOSPAQASGNQITLCISTKNLN-----TQAMV 703

Db 663 EQAIKOVNAKQ-----ALNGNANVQHADEATALINSSNDLNAQKDALKQOV 711

Qy 704 ENGGLTIVGGI-----YEDNGNTLTVPILG--DIPVGNLFKTRGKKT-----747

Db 712 QNATT--VAGVNVVKQTAQELNNAMTQLKQGIADKEQTKADGNFVNADPKQNAVNAQAVA 769

Qy 748 RRELLIFITPRIMGT 762

Db 770 KAEALISATPDVVVT 784

RESULT 38

AAU37487

ID AAU37487 standard; Protein; 875 AA.

XX

AC AAU37487;

XX

DT 14-FEB-2002 (first entry)

XX

DE Staphylococcus aureus cellular proliferation protein #1657.

XX

KW Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

XX

OS Staphylococcus aureus.

XX

WO200170955-A2.

XX

PD 27-SEP-2001.

XX

PF 21-MAR-2001; 2001WO-US09180.

XX

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX

PA (ELIT-) ELITRA PHARM INC.

XX

PI Haselbeck R, Onlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX

DR WPI; 2001-611495/70.

DR N-PSDB; AAS55346.

XX

New polynucleotides for the identification and development of

antibiotics, comprise sequences of antisense nucleic acids -

XX

Example 3; Seq ID No 13080; Sllpp; English.

XX

The invention relates to antisense inhibitors of genes essential to

prokaryotic cellular proliferation, their use in identifying the

genes, their use in the discovery of novel antibiotics, the essential

genes themselves and the encoded proteins. The prokaryotes used are

Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

invention is also useful for the identification of potential new targets

for antibiotic development. The antisense nucleic acids can also be used

to identify proteins used in proliferation, to express these proteins,

and to obtain antibodies capable of binding to the expressed proteins.

The proteins can be used to screen compounds in rational drug discovery

programmes. The antisense nucleic acid sequence is also useful to screen

for homologous nucleic acids which are required for cell proliferation in

a wide variety of organisms. The present sequence represents an

essential prokaryotic cellular proliferation protein.

Note: The sequence data for this patent did not form part

of the printed specification, but was obtained in electronic

format directly from WIPO at

ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 875 AA;

Query Match 4.1%; Score 159; DB 22; Length 875;

Best Local Similarity 19.1%; Pred. No. 0.0072;

Matches 175; Conservative 129; Mismatches 303; Indels 308; Gaps 42;

Qy 1 MNTKLTIKIISGLFVATAAFQTASAGNITDIKVSSLPNKQKIV--KVSFDKILVNPFGVFT 58

Db 62 LNTAMTALKRA--IADKA-ETKASGNY---VNADANKRQAYDEKVTAAENIVSGTPTPT 114

Qy 59 SSPARIALDFEOTGISMDQOVLEVADPLLSKISAQNSSRARLVNLNPKPOYNTEVRGN 118

Db 115 LTPADVT-----NAATQVNTAKTQLN-----GNHNLVAKQ 145

Qy 119 KVMIFINESDD---TVSAPARPAVKAAPAKQOG-CRTVYQVRSIR-----IOTLVPK 169

Db 146 NA-----NTAIDGLTSLNGPQKAKLK-----EQVGQATTLPNVQTVRDNAQTLLNTAMKG 194

Qy 170 KTTAAAPTESVVSUSAPSPAKO-----QAAASAKO---QTAAPA-KQOTAPAPAKQ 218

Db 195 LEDSLA--NEATIKAGQNYTDASQNKQTDYNSAVTAAKAIQTITSPSNQAEINQAKQ 252

Qy 219 AAAPAKQTNIDPRKDGKNAGIIEAALGFAGQPDISOQHDHIIIVTLKNHTLPTTLQRLSD 278

Db 253 VTAKQOALN-----GQENL-----RTAQNAKQHLN 278

Qy 279 VADFTFVQKVLTKR-----LNNDTQLIITTAGNWLNVKSAAPGYFTFOVLKKQNL 331

Db 279 GLSDLTDAQKDAVKRQIEGATHVNEVTQ---AQNADALNTAMT-----NL 321

Qy 332 ESG-----GVNNAKPTFTGRKISLDFQDVE-----RTILQILAKESGMNI 372

Db 322 KNGIQDQNTIKQGVN-----FTDDEAKRNAYTNVTAQEIILNKAQGN- 366

Qy 373 VASDSVNGKRWLSLKDVPWDQALDVMQARNLDMQOQGNVNIAPRDELLAKDAFLQAE 432

Db 367 TSKDGV-----ETALENVQRAKN-ELNGNQNVAN-----AKTTAK 400

Qy 433 KOIADIGALYSQNFQLYKYNVEEFSILRLDNADTTGNR-NTLVSGRGSVLIDPATNTLI 491

Db 401 NALNLTSSINNAQKEALKSQIEGATTVAGVNVQSTTASELNTAMNLQNGINDEATKAA 460

Qy 492 VTDRSV-----IEKFRKLIDELDVPAQOVMI EARIVEAADGFSRDL-----533

Db 461 LNGTQNLERAKQHANTAIDGLSHLTNAQKEALKQLVQQSTTVAAEQNEQKANNVDAAMD 520

Qy 534 ----GVKFGATGKKK-----LKNDTSAFGWGVSFGGDDKQGAETKINLPITAAANSI 583

Db 521 KURQSIADNATTKQONQNYTDAQNKDKDAYNNAVTTAQGIIDQTTSTLDPVTINQAGOV 580

Qy 584 SLVRAISSGALNLELS---ASELSKTKTLANPRVLT-----QNRK- 621

Db 581 SITKNALNGNENLEAAKQQAQSQSLGSLDLNNAQKQTVTDQINGAHTVDEANQIKQNAQN 640

Qy 622 -----BAKIESGYEIPFT-----VTSIANGSSSTNTEL 649

Db 641 LNTAMGNLQAIADKADKATKATVNTFTDADQAK-QQAYNTATVNAENIISKANGNATQAEV 699

Qy 650 KKAVLGLTVPITPDGQIIMTVKINKDSPACASGNQITLCISTKNLN-----TQAMV 703

Db 700 EQAIKOVNAKQ-----ALNGNANVQHADEATALINSSNDLNAQKDALKQOV 748

Qy 704 ENGGLTIVGGI-----YEDNGNTLTVPILG--DIPVGNLFKTRGKKT-----747

Db 749 QNATT--VAGVNVVKQTAQELNNAMTQLKQGIADKEQTKADGNFVNADPKQNAVNAQAVA 806

Qy 748 RRELLIFITPRIMGT 762

Db 807 KAEALISATPDVVVT 821

RESULT 39

AAU34339







```

564 QY DKWGAETKINLPITAAANSISILVRAISSGALNLELSASBSLSKTKTTLANPRVLUTQNRKEA 623
      :      :      :      :      :      :      :      :      :      :      :
717 Db NGGNADLK-NLNVHAKSGALNI--HSDRALSTENTKLESTHNLNACHERVTLAQVDA 772
      :      :      :      :      :      :      :      :      :      :      :
624 QY -----KTESYEVPFTVTSANG-----GSSTNTELKKAVLGTVTPNIT 663
      :      :      :      :      :      :      :      :      :      :      :
773 Db YAIRHLSITGSIQWQNDKLPANKLVANGVLNARYSQIADNTTLRAGAINLTAGTALV 832
      :      :      :      :      :      :      :      :      :      :      :
664 QY PDGQI---IMTVKINKDSP-----AQCASGNQITLCISTKNLNTQAMVENGTTLI 710
      :      :      :      :      :      :      :      :      :      :      :
833 Db KRGINNKSTVSTKTLEDNAELKPLAGRLNIYEAAGSTLTIEPANRISAHTDLSIKTGCKLL 892
      :      :      :      :      :      :      :      :      :      :      :
711 QY VGGIYEEDNGNTLTKVPLLDGPVIGNLFXTRGCKKTDRELLIFITPRIMGTAGNSL 767
      :      :      :      :      :      :      :      :      :      :      :
893 Db LSA----KGGNAGAPSAQVSLEAKGNIRLVGTG-ETDLRGSKI-----TAGKNL 936
      :      :      :      :      :      :      :      :      :      :      :

```

Search completed: December 9, 2003, 10:25:47  
Job time : 55 secs

Grandi G, Masignani V, Pizzo M, Rappuoli R, Scarlato V;  
WPI; 1999-444400/37.  
N-PSDB; AAX99152.  
New protein and its nucleotide sequence, useful in vaccines or  
diagnostic compositions for treating and/or preventing *Neisseria*  
meningitidis infections  
Claim 4; Page 88; 123pp; English.  
The invention provides proteins (AAY27201-245) from *Neisseria*  
meningitidis (strains A and B) and nucleic acid sequences (AAX99123-167)  
encoding the proteins. Compositions comprising the protein, nucleic acid  
or antibody specific to the protein are useful as pharmaceuticals, e.g. a  
vaccine composition or a diagnostic composition. The composition is also  
useful for treating or preventing an infection due to *Neisseria*  
bacteria, especially *Neisseria meningitidis*.  
Sequence 1978 AA;  
Query Match 4.1%; Score 156.5; DB 20; Length 1978;  
Best Local Similarity 20.2%; Pred No. 0.037;  
Matches 193; Conservative 133; Mismatches 370; Indels 261; Gaps 44;  
QY 1 MNTKLTIIISGLFVATAFOTASAGNTIDIKVSLPNKKIV---KVGFDEIVN---PTG 55  
DB 51 LKTTLKTLCVSL-VSLSWLPAHAQITTD---KSAKKNQQVILKTTGAPLVNIQTENG 106  
QY 56 FVTSSPARIALDPQTGISDQOVLEVADPULLKISAAQNSSRARLVNLNKPQYNTVE 115  
DB 107 RGLSHNRYTQFDVNDKGAVLNND--RNNPFVWGS-----AQLLIN-----EV 148  
QY 116 RGKN-----VWIFINESDDTVSAPAPVKA-----APAPA----- 147  
DB 149 RGTASKLNGIVTGGQKADVIIANPNTVNGGGFKVGRGILTTGAPOIGKDGALTGFD 208  
QY 148 KQGCRTV-----YQVRSIRIOTLYPGKTTAAAPFTESVWSVSAFFSPAK 192  
DB 209 VRQGTLTVGAAWMDKGGADVTGVLAVALVGLKQKLNLAIVTGPQKVDYVASEISAG- 267  
QY 193 QQAASAKQOTAAAPAKQOATAAPAKQAAAPAKOTNIDFRKG---KNAGIIELA---ALG 246  
DB 268 -----TAAGTKPIALDTRALGGMYADSIILIANXGVGVKNAGTLEAAKQLIVT 317  
QY 247 FAGQPDISOQHHIIVTKNHTLPTTLORSLDVAD-----FKTFVQVATLKLNNDTQL 300  
DB 318 SSGRIENSGR---IAITADGTEASPTVLSIETTEKGAAGTFISNGGRIESKGL---L 368  
QY 301 IITTAGNWEL-----VNKSAAPGVFTFQ-----VLPKKQNLSESGVNNNA---PKTFT--G 345  
DB 369 VIETGEDISLNGAVVQNGSRPATTVLNAGHNLVIESKTN-----VNNAKGPATLSADG 423  
QY 346 RKISLDFQOVEIRTILOILAKESG-----MNVIASDSVNGKMTLSLKVDVPW 391  
DB 424 RTV-----IKEASITQTTTVYSSKNGNAELGNNTRIITGADVTLSNGTISSESAVIDAKDTAH 480  
QY 392 DOA---LDLVMOQRNLDMROQNI VNIAPREDLLAKRAFLQAEKDIAADLGALYEQNFOL 448  
DB 481 IEAGKPLSEASTVTSIDRLNGKSIKGGKQALALLADDNITAKT-TNLNTPGNLYVHTGKD 539  
QY 449 KYKNVEEFSI---LRLDN-ADTTGNRLTVSGR-----GSVLIDPATNLIVDTFRS 497  
DB 540 LNLNVYDKLSAASIHKLSDNAAHIITGTSKTLTASKDMGVVEAGSLNV---TNTNLTNRSNG 596  
QY 498 VIEKFRK-----LIDELDVPAQO--VMIEARIVEAADG----- 528  
DB 597 LHIQAAKGNILQNLTKLNAAKALETITALOGNIVSDGLHAVSDAGHVSULLANGNADFTGHN 656  
QY 529 -FSRDLGVKFGATGKKLKNQ-----TSAPFGWGVNSGFG-----GD 563  
DB 657 TLITAKADYNAGSVGKGRLLKADNTNITSSSGDITLVAGNGIQLGDKGQRNSINGRHSIKN 716

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 10:27:18 ; Search time 36 Seconds

(without alignments)  
3972.818 Million cell updates/sec

Title: US-09-701-271A-2

Perfect score: 3848  
Sequence: 1 MNTKTKIISGLFVATAAFQ.....ELLIFITPRINGTAGNSLRY 769

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Data base : Published Applications\_AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description         |
|------------|-------|-------------|--------|----|---------------------|
| 1          | 3707  | 96.3        | 769    | 12 | US-10-320-800-10    |
| 2          | 332   | 8.6         | 654    | 9  | US-09-815-242-10361 |
| 3          | 300   | 7.8         | 776    | 9  | US-09-815-242-5092  |
| 4          | 167   | 4.3         | 1029   | 9  | US-09-815-242-5885  |
| 5          | 167   | 4.3         | 1048   | 9  | US-09-815-242-13083 |
| 6          | 166.5 | 4.3         | 2086   | 9  | US-09-815-242-5639  |
| 7          | 166.5 | 4.3         | 2659   | 12 | US-10-311-879-28    |
| 8          | 166.5 | 4.3         | 5795   | 9  | US-09-815-242-12610 |
| 9          | 160.5 | 4.2         | 1073   | 12 | US-10-193-764-45    |
| 10         | 160.5 | 4.2         | 1079   | 12 | US-10-193-764-43    |
| 11         | 159.5 | 4.1         | 6281   | 9  | US-09-815-242-12996 |
| 12         | 159   | 4.1         | 837    | 9  | US-09-815-242-5883  |
| 13         | 159   | 4.1         | 875    | 9  | US-09-815-242-13080 |
| 14         | 158   | 4.1         | 2434   | 9  | US-09-815-242-5935  |
| 15         | 155.5 | 4.0         | 1095   | 12 | US-10-193-764-55    |

|    |       |     |      |    |                     |                   |
|----|-------|-----|------|----|---------------------|-------------------|
| 16 | 155.5 | 4.0 | 1536 | 12 | US-10-193-764-63    | Sequence 63, Appl |
| 17 | 155.5 | 4.0 | 1536 | 14 | US-10-092-880-2     | Sequence 2, Appl  |
| 18 | 155.5 | 4.0 | 3241 | 9  | US-09-841-786-1     | Sequence 1, Appl  |
| 19 | 150.5 | 3.9 | 714  | 9  | US-09-841-786-4     | Sequence 4, Appl  |
| 20 | 150.5 | 3.9 | 2368 | 9  | US-09-815-242-5635  | Sequence 5635, Ap |
| 21 | 150.5 | 3.9 | 2368 | 9  | US-09-815-242-12389 | Sequence 12389, A |
| 22 | 150   | 3.9 | 1599 | 14 | US-10-092-880-9     | Sequence 9, Appl  |
| 23 | 147.5 | 3.8 | 1838 | 9  | US-09-879-248-8     | Sequence 8, Appl  |
| 24 | 147.5 | 3.8 | 1981 | 10 | US-09-928-457-38    | Sequence 38, Appl |
| 25 | 147.5 | 3.8 | 2015 | 15 | US-10-066-551-1     | Sequence 1, Appl  |
| 26 | 147.5 | 3.8 | 2478 | 9  | US-09-815-242-5816  | Sequence 5816, Ap |
| 27 | 147.5 | 3.8 | 2478 | 9  | US-09-815-242-12967 | Sequence 12967, A |
| 28 | 146.5 | 3.8 | 1180 | 12 | US-10-193-764-61    | Sequence 61, Appl |
| 29 | 146.5 | 3.8 | 1188 | 12 | US-10-193-764-59    | Sequence 59, Appl |
| 30 | 145   | 3.8 | 1741 | 10 | US-09-971-536-68    | Sequence 68, Appl |
| 31 | 143.5 | 3.7 | 1220 | 12 | US-10-193-764-28    | Sequence 28, Appl |
| 32 | 143.5 | 3.7 | 1226 | 12 | US-10-193-764-26    | Sequence 26, Appl |
| 33 | 143   | 3.7 | 1600 | 14 | US-10-092-880-10    | Sequence 10, Appl |
| 34 | 143   | 3.7 | 2076 | 9  | US-09-815-242-5815  | Sequence 5815, Ap |
| 35 | 143   | 3.7 | 2186 | 9  | US-09-815-242-12913 | Sequence 12913, A |
| 36 | 142.5 | 3.7 | 915  | 12 | US-10-193-764-35    | Sequence 35, Appl |
| 37 | 141.5 | 3.7 | 1222 | 12 | US-10-193-764-37    | Sequence 37, Appl |
| 38 | 141.5 | 3.7 | 1228 | 12 | US-10-193-764-34    | Sequence 34, Appl |
| 39 | 141   | 3.7 | 2353 | 10 | US-09-797-862-33    | Sequence 33, Appl |
| 40 | 138   | 3.6 | 992  | 12 | US-10-193-764-57    | Sequence 57, Appl |
| 41 | 138   | 3.6 | 998  | 12 | US-10-193-764-55    | Sequence 55, Appl |
| 42 | 138   | 3.6 | 2514 | 12 | US-10-320-800-40    | Sequence 40, Appl |
| 43 | 137   | 3.6 | 2834 | 12 | US-10-085-959-252   | Sequence 252, App |
| 44 | 135.5 | 3.5 | 2189 | 12 | US-10-172-502-2     | Sequence 2, Appl  |
| 45 | 134.5 | 3.5 | 3788 | 11 | US-09-952-267-76    | Sequence 76, Appl |

ALIGNMENTS

RESULT 1  
US-10-320-800-10  
; Sequence 10, Application US/10320800  
; Publication No. US20030215469A1  
; GENERAL INFORMATION:  
; APPLICANT: ROBINSON, ANDREW  
; APPLICANT: GORRINGE, ANDREW  
; APPLICANT: HUDSON, MICHAEL  
; APPLICANT: REDDIN, KAREN  
; TITLE OF INVENTION: MULTICOMPONENT MENINGOCOCCAL VACCINE  
; FILE REFERENCE: 1581, 0790001  
; CURRENT APPLICATION NUMBER: US/10/320,800  
; CURRENT FILING DATE: 2002-12-17  
; PRIOR APPLICATION NUMBER: PCT/GB99/03626  
; PRIOR FILING DATE: 1999-11-02  
; NUMBER OF SEQ ID NOS: 75  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 769  
; TYPE: PRT  
; ORGANISM: Neisseria meningitidis  
US-10-320-800-10

|                       |                 |  |           |             |
|-----------------------|-----------------|--|-----------|-------------|
| Query Match           | 96.3%;          | Score 3707;  | DB 12;    | Length 769; |
| Best Local Similarity | 96.9%;          | Pred. No. 2e-291;  |           |             |
| Matches 745;          | Conservative 3; | Mismatches 21;   | Indels 0; | Gaps 0;     |
| Qy                    | 1               | MNTKTKIISGLFVATAAFQTASAGNITDKVSSLNPKQIKVVSFDEKIVNPTGVTSS   | 60        |             |
| Db                    | 1               | MNTKTKIISGLFVATAAFQTASAGNITDKVSSLNPKQIKVVSFDEKIVNPTGVTSS   | 60        |             |
| Qy                    | 61              | PAIADDFEQTGISMQQOVLVADPLLSKISAQNSRRARLVNLNPKQYNTVEGRNKV    | 120       |             |
| Db                    | 61              | PAIADDFEQTGISMQQOVLVADPLLSKISAQNSRRARLVNLNPKQYNTVEGRNKV    | 120       |             |
| Qy                    | 121             | WIFINESDDTVSAPAPAVKAAAPAAKQGGCTVTVQVRSIRIQTLYPKKTTAAAPFTES | 180       |             |
| Db                    | 121             | WIFINESDDTVSAPAPAVKAAAPAAKQGGCTVTVQVRSIRIQTLYPKKTTAAAPFTES | 180       |             |

```
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10361
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-815-242-10361

Query Match      8.6%; Score 332; DB 9; Length 654;
Best Local Similarity 21.5%; Pred. No. 5.2e-18;
Matches 123; Conservative 97; Mismatches 183; Indels 168; Gaps 17;

Qy 181 VVVSAPESPAAQAAAQAAKQOATAAPAKQAAAPAKQTNIDFRKDGKAGII 240
Db 181 VVVSAPESPAAQAAAQAAKQOATAAPAKQAAAPAKQTNIDFRKDGKAGII 240
Qy 241 ELAALGFAGQDPSOQHIIIVTLKNHTLPTTLQSSLDVADFKTPQVKVTLKRLNNDTOL 300
Db 241 ELAALGFAGQDPSOQHIIIVTLKNHTLPTTLQSSLDVADFKTPQVKVTLKRLNNDTOL 300
Qy 301 IITTAGNMELNKSAAQGYFTFQVLPKKQNLSESGVNNAPKTFTRKISLDFQDVEIRTI 360
Db 301 IITTAGNMELNKSAAQGYFTFQVLPKKQNLSESGVNNAPKTFTRKISLDFQDVEIRTI 360
Qy 361 LOILAKESGMNIVASDSVNGKMTLSLKDVPMDQALDLVQARNLDMROQGNVNIAPRDE 420
Db 361 LOILAKESGMNIVASDSVNGKMTLSLKDVPMDQALDLVQARNLDMROQGNVNIAPRDE 420
Qy 421 LIAKDKAFIQAEKDIAADLGALYSQNFQKLYKNVEEFRSILRLDNADITGNENTLVSGRGS 480
Db 421 LIAKDKAFIQAEKDIAADLGALYSQNFQKLYKNVEEFRSILRLDNADITGNENTLVSGRGS 480
Qy 481 VLIDPATNTLIVTDRSVIEKFKLIDELDPVPAQOVMEARIVEAADGFSRDLGVKFGAT 540
Db 481 VLIDPATNTLIVTDRSVIEKFKLIDELDPVPAQOVMEARIVEAADGFSRDLGVKFGAT 540
Qy 541 GKXKLKNDTSAPFGWNSGFGGDDKWAETKINLPITAAANSISLVRAISSGALNLELSA 600
Db 541 GKXKLKNDTSAPFGWNSGFGGDDKWAETKINLPITAAANSISLVRAISSGALNLELSA 600
Qy 601 SESLSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSIANGSGSNTNTELKXAVLGLTVTP 660
Db 601 SESLSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSIANGSGSNTNTELKXAVLGLTVTP 660
Qy 661 NITPGQIIMTVKINKDSPAQACASQNTILCISTKNLNTQAMVNGGTLIVGGIYEEDNG 720
Db 661 NITPGQIIMTVKINKDSPAQACASQNTILCISTKNLNTQAMVNGGTLIVGGIYEEDNG 720
Qy 721 NLTLPVLLGDIPIVGNLFKTRGKKTDRRELLIFITPRIMGTAGNSURY 769
Db 721 NLTLPVLLGDIPIVGNLFKTRGKKTDRRELLIFITPRIMGTAGNSURY 769

RESULT 2
US-09-815-242-10361
; Sequence 10361, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
```

```
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10361
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-815-242-10361

Query Match      8.6%; Score 332; DB 9; Length 654;
Best Local Similarity 21.5%; Pred. No. 5.2e-18;
Matches 123; Conservative 97; Mismatches 183; Indels 168; Gaps 17;

Qy 351 DPQVIRITLOILAKESGMNIVASDSVNGKMT-----LSLKD----- 388
Db 351 NFNADIRQFVEIVGQHLKLTILIDPSVQGTISVRNSDFTSQOQYVFFLSLIDLGYSV 94
Qy 389 -----VPWDQ-----ALDLVQARN-LDM 406
Db 95 IITLNDGFLKVVRSANVKTSPGMADSSRPGVGDDELVTRIVPLENVFARDLAPLRQMMDA 154
Qy 407 RQGNINVINIAPRDELLAKOKA-----FLQAEKDIAADLGALYSQNFQKLYKNVEEFRSILR 461
Db 155 GSVGNVHVHPEPSNVILITGRASTINKLIEIKVVDVIGTEKQOIILHLEYASAEADLAEILN 214
Qy 462 -----LONADITGNRNT-----LVSG----- 477
Db 215 QLISESHGSKSQMPALLSAXIVADKRTNSLIISOPEKARQRTSLLSLDVSESEEGNTRV 274
Qy 478 -----RG-----SVLIDPATNTLIVTDRSV 498
Db 275 YLYKAYAKATNLVLEVTGVSSEKLDKGNARKPKSSGAMDNVAITADEQTNLSLVIDAQSV 334
Qy 499 IEKFKLIDELDPVPAQOVMEARIVEAADGFSRDLGVKFG-----ATGKKLKNDSAFSGWGV 556
Db 335 QEKLATVIARLDITRRACQVLEAIIVEVDGNGLNLGVQWANKVNGAQQTNT-TGLPIFNA 393
Qy 557 NSGFGGDDKWAETKINLPITAAANSISLVRAISSGALN-----LELSASESLSKTKTLAN 612
Db 394 AQGVADYKXNGGITSAN-----PAWDMFSAYNGWAAGFFNGDGVLLTALASNNKNDILAT 449
Qy 613 PRVLTQNRKEAKIESGYEIPFTVTSIANGSGSNTNTELKXAVLG-----LTVTPNITPDQIIM 670
Db 450 PSIVTLNDKLASFNVGQDVPVLSGQSTTSGDNVNTVERKTVGTLKLVTPQVN-EGDAVL 508
Qy 671 TVKINKDSPAQACASQNTI-LCISTKNLNTQAMVNGGTLIVGGIYEEDNGNLTLPVLL 729
Db 509 -LEIEQVSSVSDSSNSTLGPFTFTRTIQNAVILVKTGETVVLGGLDDFSKEQVSKVPLL 567
Qy 730 GDIPVIGNLFKTRGKKTDRRELLIFITPRIM 760
Db 568 GDIPVIGLQFRYTSRTERAKNLMVFIRPTII 598

RESULT 3
US-09-815-242-5092
; Sequence 5092, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
```

```

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-32
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5092
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-08-815-242-5092

```

| Query Match               | 7.8t; Score 300; DB 9; Length 776;                                |
|---------------------------|---|
| Best Local Similarity     | 21.4t; Pred. No. 2.7e-15;   |
| Matches 122; Conservative | 97; Mismatches 180; Indels 170; Gaps 16;                          |
| 349 QY                    | SLDFQDVETRIQILAKESGMNIVASDSVNGKMT---LSLKDVPWDQALDLVWQAR 402       |
| 47 Ddb                    | TKNMKDAEIGDFIEQVSSISGQTFVVDPRVKGRVTVVSQARLSLAEV--YQLFLSVLATH 104  |
| 403 QY                    | NLDMRQOGNIVNIATPRODELLAKOKAFLOAEKDIAD--- 437                      |
| 105 Ddb                   | GYAVLPDQDQARIVENWE--ARQCA---AQKTVRDGFGSLRVTVVQAQOTSVAELIPMR 159   |
| 438 QY                    | LGALYSQNQL---KYKNVEFFRSILR----- 461                               |
| 160 Ddb                   | PLVPAGHGLAAVPSANALIVSDRRSNIETIEAIVRSLSRAGEHDYSIVDMRHAWVAEIAE 219  |
| 462 QY                    | LDN-----ADTTGNRNTLV----- 475                                      |
| 220 Ddb                   | VLDSVTPAAKSNATVQVVIADSRNRLVLGFPQPARLRLLAQSLDVPSSRSANSRVI 279      |
| 476 QY                    | -----SGRGS-----VLIDPATNLIVTDSVTEKF 502                            |
| 280 Ddb                   | RLRHGDAKTLAATLGEIGESLHGERGQDQSGKRGLLVRADESLNALVILADPEDVGLL 339    |
| 503 QY                    | RKLDELDPQAQVMIEARIVEAADGFSRDLGVKFGATGKKLKNDTSAFCMG--VNSGFG 561    |
| 340 Ddb                   | EDIVRLQDLVPAQLLVEAAIVELSGEIGDALGVQWA-----LRSHVAGAGAGADSGLS 393    |
| 562 QY                    | GDDKWGAETKINIPITAAANSISLVRASIGALINLELSASELSKTKTLANPRVLTQNRK 621   |
| 394 Ddb                   | IGTLLGALQAGKPAELPDGAIVLGSRDFGALVTALSRN---SRNSNLSTPSLTLDNQ 450     |
| 622 QY                    | EAKIESGYEIFF---VTVSTANGSGSNTTELKKAVALGLT--VTFNITPDGQIIMTV--KI 674 |
| 451 Ddb                   | KAEILVGNQVFPFGTSYTTASGSSNPFTTVERKDIGVTLKVTPHIGEDRMRLREIQEI 510    |
| 675 QY                    | NKDSPAQACAGNQTILCISFNLTQAMVENGSTLIVGIEEONGNTLTKVPLIGIPV 734       |
| 511 Ddb                   | SSIAPTATLAAKAVDLVTNKRSTKSTVLADDCQVIVGLGIQDDLQSRSDSRVPLDIPG 570    |
| 735 QY                    | IGNLFTKRGKKTDRRELLIPITPRMGTA 763                                  |
| 571 Ddb                   | VGRLFSSRRETRVKNRLMWFLRESIVRDA 599                                 |

RESULT 4  
US-09-815-242-5885  
; Sequence 5885, Application US/09815242  
; Patent No. US2020061589A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Orlsen, Kari L.

```

1  APPLICANT: Zykind, Judith W.
2  APPLICANT: Wall, Daniel
3  APPLICANT: Trawick, John D.
4  APPLICANT: Carr, Grant J.
5  APPLICANT: Yamamoto, Robert T.
6  APPLICANT: Xu, H. Howard
7  TITLE OF INVENTION: Identification of Essential Genes in
8  TITLE OF INVENTION: Prokaryotes
9  FILE REFERENCE: ELITRA.011A
10 CURRENT APPLICATION NUMBER: US/09/815,242
11 CURRENT FILING DATE: 2001-03-21
12 PRIOR APPLICATION NUMBER: 60/191,078
13 PRIOR FILING DATE: 2000-03-21
14 PRIOR APPLICATION NUMBER: 60/206,848
15 PRIOR FILING DATE: 2000-05-23
16 PRIOR APPLICATION NUMBER: 60/207,727
17 PRIOR FILING DATE: 2000-05-26
18 PRIOR APPLICATION NUMBER: 60/242,578
19 PRIOR FILING DATE: 2000-10-23
20 PRIOR APPLICATION NUMBER: 60/253,625
21 PRIOR FILING DATE: 2000-11-27
22 PRIOR APPLICATION NUMBER: 60/257,931
23 PRIOR FILING DATE: 2000-12-22
24 PRIOR APPLICATION NUMBER: 60/269,308
25 PRIOR FILING DATE: 2001-02-16
26 NUMBER OF SEQ ID NOS: 14110
27 SOFTWARE: FastSeq for Windows Version 4.0
28 SEQ ID NO 5885
29 LENGTH: 1029
30 TYPE: prt
31 ORGANISM: Staphylococcus aureus
32 US-09-815-242-5885

```

|     |  |                   |                    |             |              |
|-----|--|-------------------|--------------------|-------------|--------------|
|     | Query Match  | 4.3%              | Score 167;         | DB 9;       | Length 1029; |
|     | Best Local Similarity  | 19.6%;            | Pred. No. 0.00025; |             |              |
|     | Matches 146;   | Conservative 112; | Mismatches 295;    | Indels 192; | Gaps 30;     |
| 81  | EYADPILLSKIGAAQNS-SRARLVNLNKPQQYNTEVRGNKWIFINESDDTVTSAPARPAV   | 139               | :                  | :           | :            |
| 223 | DYVDANDXQTAYNNVAATAETIINANSNPENST-----ITQAEQVNS-SKTAL          | 273               | :                  | :           | :            |
| 140 | KAAPAAPAQOGCCTVYGVSIRIQTLYPEGKTTAAAPTESVWS--VASPFSPAKQAAAA     | 197               | :                  | :           | :            |
| 274 | NGDENLTPAAQNAKTY-----LNLT---TSDAQKNLLISOITSATRVSGVDVTVKQ       | 322               | :                  | :           | :            |
| 198 | SAGOOTAAPAKOQTAAPAKQAQAAAAPAKOTNIDFRKGKNVAGIIELAAJGFACQPDISQH  | 257               | :                  | :           | :            |
| 323 | NACHLDQWASLONGINNSESQKSSEKYRDADTNK-----QQEY                    | 361               | :                  | :           | :            |
| 258 | DHIIVT----LKNHLPPTLQRSILDVADFKEPVOKVILKR--LNNDTOLI             | 301               | :                  | :           | :            |
| 362 | DNAITAIAAILINKSTGPNTAQNAVEAA-----LRVNNAKDALNGDAKLIAAQNAAKOHL   | 416               | :                  | :           | :            |
| 302 | -----ITTAGNWELVNK-SAAPGYFTFOVLPEKKONLES-----GGVNNAPKTFTRCKISL  | 350               | :                  | :           | :            |
| 417 | GTLTHITTAQRNDLTNQISOATNLAVESV--KQANSLDGMGNLOTAINDKSGTLASQ      | 474               | :                  | :           | :            |
| 351 | DFQDVE-----IRTIOILIAKSSGMNVIASDSVNGKRTLSLKDVPMDQALDLVNYQ       | 400               | :                  | :           | :            |
| 475 | NFLDADEQRKNAYNQAVSAABETILNKGTGPN-TAKTAV-----EQALLNVNN          | 520               | :                  | :           | :            |
| 401 | AR-----NLDMEQGQGVNIPIARPDELLAKDKAFLOAEKIADLGALYSQNFOLKYKVY     | 453               | :                  | :           | :            |
| 521 | AKHALNGTONLNNAKQAAITANGASGLDNQOKDALKAQAN---GAORVSNQAODVQHNA    | 576               | :                  | :           | :            |
| 454 | BEPFSIL-----RLNDAOTT-----GNRNTLYSVGRGSVLIDPAT                  | 487               | :                  | :           | :            |
| 577 | TCLMTANGTLKHAITADKNTLTASSKYVNADETQKNAYTTTKVTNAEHIIISGTTVVTTFSE | 636               | :                  | :           | :            |
| 488 | NTLIVTDRSV-----IEFKRK-----LIDEJDVPAQQVMIBEARIVEA-              | 525               | :                  | :           | :            |
| 637 | VTRAANOVSNAKBQLNGDERLREAKONANTAI DALTOLTPT-OKAKUKEOVGOANRLBDV  | 695               | :                  | :           | :            |

QY 526 -----ADGFSRDLGVKFGATGKKLKNDSAFGWGNSGFGGDDKWA 568  
Db 696 QTVQNGQALNNAMKGLRDSIANETTVKTSQNYVDASPNQSTYNSAVSNAKGIINQNN 755  
QY 569 ETKINLPITAAANSISLVRAISSGALNLSASELSKTKTLANPRVLTQNRKEA---KI 625  
Db 756 PTMDTSAITQATQVNNAKNGLGAENLRNAQNTAKQNLTLSH---LTNNKSAISSQI 812  
QY 626 ESGYEIPFTVTSIANGSSSTNTLKKAVLGLTVTNITPDGQIIMTVKINKDSPACQASG 685  
Db 813 DRAGHVS-EVTATKNAATELNTQMGLEQAIHQNTVKQSVKFTDADKAKRDAYTNAVR 871  
QY 686 NQTILCISTKNLNT-----QAMVEN 705  
Db 872 ABAIL-NKTQAGANTSKQDVEAAIQN 895

RESULT 5  
US-09-815-242-13083  
; Sequence 5639, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; NUMBER OF SEQ ID NOS: 1410  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13083  
; LENGTH: 1048  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus

US-09-815-242-13083

Query Match 4.3%; Score 167; DB 9; Length 1048;  
Best Local Similarity 19.8%; Pred. No. 0.00026;  
Matches 146; Conservative 112; Mismatches 295; Indels 192; Gaps 30;

QY 81 EVADPLLSKISAQNS-SBARLVNLNKGQVNTVRGNKWFNFINESDDTVSAPAPAV 139  
Db 235 DYVDANDQQTAVNNAVAEAETIINANSNPKNPST-----ITQKAEQVNS-SKTAL 285  
QY 140 KAAPAAPAKQCGRTVYQVRSRIQTLYPGKTTAAAPFTESVVS---VSAPSPAKQAAAA 197  
Db 286 NGDENJTAARQNAKTY-----LNTL-----TSITDAQKNLISQITSATRVSGVDTVKQ 334  
QY 198 SAKQQTAAAPAKQQTAAAPAKQQAAPAKQNTIDFRKDGKNGAGIIEAALGFAQGPDISQOH 257  
Db 335 NAQHLDAQANASLQNGINNSQVKSSEKYDADTNK-----QOEY 373

QY 258 DHIIVT-----LKNHTLPTTLQSRSLDVADFPTVQKVTLKE--LNNDTQLI----- 301  
Db 374 DNAITAAKAILNKSTGNTAONAVEA-----LQRVNNAKDALNGDAKLIAAQAQNAKHOHL 428  
QY 302 -----ITTAGNWLNVK-SAPGVYFTFOVLPKKNLLES-----GGVNNAPKTTTGRKISL 350  
Db 429 GTLTHITTAORNDLTNISOATNLNLAGVESV--KONANSLDGAAGNLTQATINDSGSLTASQ 486  
QY 351 DFQDVE-----IRTLQILAKESGMNIVASDSVNGKMTLSLKDVDPWDQALDLVMQ 400  
Db 487 NFLDADSEKRNAYNAOVAASAAETILNKOTGPN-TAKTAV-----EQALNNVNN 532  
QY 401 AR-----NLDMRQGNIVNIAPRODELLAKDAFLQAEKDIAADLGLALYSONFOLKYKNV 453  
Db 533 AKHALNGTONLNNAKQAAITAINGASDLNOKQDKALKAQAN-----GAQSVNSAODVOHNA 588  
QY 454 EEFRSIL-----RLDNADTT-----GNRNTLVSGRGSVLIDPAT 487  
Db 589 TELNTAMGTLKHAIDAKTNTLASSKYVNAADSTKONAVTTKVTAENHIISGTPVVTTPSE 648  
QY 488 NTLIVTDRSV-----IEFKRK-----LIDELVPAQOVMIERAVEA----- 525  
Db 649 VTAANQVNSAKOELNGDERLREAKQKNANTAI DALTQNLTP-QKAKLKEQVGOANRLDEV 707  
QY 526 -----ADGFSRDLGVKFGATGKKLKNDSAFGWGNSGFGGDDKWA 568  
Db 708 QTVQNGQALNNAMKGLRDSIANETTVKTSQNYVDASPNQSTYNSAVSNAKGIINQNN 767  
QY 569 ETKINLPITAAANSISLVRAISSGALNLSASELSKTKTLANPRVLTQNRKEA---KI 625  
Db 768 PTMDTSAITQATQVNNAKNGLGAENLRNAQNTAKQNLTLSH---LTNNKSAISSQI 824  
QY 626 ESGYEIPFTVTSIANGSSSTNTLKKAVLGLTVTNITPDGQIIMTVKINKDSPACQASG 685  
Db 825 DRAGHVS-EVTATKNAATELNTQMGLEQAIHQNTVKQSVKFTDADKAKRDAYTNAVR 883  
QY 686 NQTILCISTKNLNT-----QAMVEN 705  
Db 884 ABAIL-NKTQAGANTSKQDVEAAIQN 907

RESULT 6  
US-09-815-242-5639  
; Sequence 5639, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16

```

; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5639
; LENGTH: 2086
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5639

Query Match          4.3%; Score 166.5; DB 9; Length 2086;
Best Local Similarity 19.9%; Pred. No. 0.00086;
Matches 175; Conservative 117; Mismatches 324; Indels 265; Gaps 37;

QY 34 SLPNKQKV-----KVSFDEKEIVNPTGFTVSSPARIALDFEQTGISMDQVLEYADPL 86
DB 110 TIANKPDVTLDAHTGKVTFTNAVTKNSAITIPK-----ACTG-----149

QY 87 LSKISAAQNSRARLVLNLNPKQGYNTEVRGNKVMFIINESDDTVSAPARPAVKAAPAP 146
DB 150 ---HGASNPS-----TLTAPATHV-----NTTEIVKDYGSNVTAEEINAVQVA---192

QY 147 AKQOGRVYQVRSIRIOTLYPGKTTAAAPT-----ESVSVSAPFSPAQQ 194
DB 193 ---NKRATATNGTAMPTNLGAGSTTIPVTVYNDGSTEEVQESI-----FTKADKR 242

QY 195 AAASAKQQTAAAP-----AKQ--TAAPAKQQA-----AAPAKQTN 227
DB 243 ELITAKNHLDDPVSDGKKPGTITQYNNAIHNAQQQINTAKTEAQQVINNERATPQQVSD 302

QY 228 IDFRKDGKNAGIELAALGAFAGQPDISQDHIIIVTLKN-----HTLPTT---LQSLD 278
DB 303 ALTKVRAAQTKINEKAL--LQNKENSO-----LVTSKNLQSSVNGVPSTGTTQOOSID 356

QY 279 VADFX-----TPVQKV-----TLKRLNNDTQLIITTAGNMELVNK--SAAPGFTFQ 323
DB 357 NYNAKKREAEITEIAAQVRVIDNGDATAQQISDEN-----TTAQRNDLTNQISOATNLAAVE 412

QY 324 VLPKQNLSEGGVNN---APKFTGRKISLDFQDVE-----IRTIQLILAKESGM 370
DB 413 SVKQANSILDGAMGLQTAINDKSGTLASQNFLDADQEKNAYNQAVSNAETILNKOTGP 472

QY 371 NIVASDSVNGKMTLSLKDVPDQALDVMQAR-----NLDNRQOQNI VNIAPRDELLA 423
DB 473 N-TAKTAV-----EQALNVNSAKHALNGTQNLNNAKQAAITAINGASDLNQ 518

QY 424 KDKAFLOAEKDIADLGALYSONFOLKYQNVBEFRSIL-----RLDNAD 466
DB 519 KOKDALKAQAN-----GAQVSNADQVORNATELNTAMGTMLKHAIADKTNTLASSKYVND 574

QY 467 TT-----GNRNLTVSGRGLVDIPATNLIVTQTRSVIEKFR-----503
DB 575 STQKNAYTKVTNAEHIISGTFVTVPFSEVTAANQVNSAKQELNGDELRVAKONANT 634

QY 504 ---KLIDELDVPAQQVMIBARIVEA-----ADGFSRDLGVKFG 538
DB 635 AIDALTQLNTP--QKAKLKEQVQANRLIEDIQVTQNGQALNNAMKGLRDSIANETTVKAS 693

QY 539 ATGKKKLKNDTSAFQGWVNSGFGGDDKWAETKINLPITAAANSILVRAISSGALNLEL 598
DB 694 QNYTDASPNNGSTYNSAVSNAKGIINQTNPTMDTSITQATQVNNAKNGLNGAENLRN 753

QY 599 SASLSKTKTLANPRVLTONRKEA---KIBSGVEIPIFTVTSIANGSSSTNTLKKAVLG 655
DB 754 AQNTAKQNLTLSH---LTNNQKSAISSQIDRAGHVS--EVTAAKNAATELNTQMGNLEQA 809

QY 656 LTVTENIIPDQGIINTVKINDSPAQCSNGQITLJCISTKNLNT-----QAMVYENGTLI 710
DB 810 IHDQNTVQGVNFTDADKAKEDATVNAVSRABETIL--NKTQANTSKODVEAATON-----863

QY 711 VGGIYEEDNGNTLTKVPLLLGDIPIVGNLFTKTRGKKTDRREL 751
DB 864 -----VTSAKNALNGQNV-----TNAKNVAKHAL 888

```

QY 648 E-----LKKAVLGLTVTNPITPDGQIIMTKVI-----NKDSP-----AQ 681  
Db 1222 QIAAEKTKVEEYNSLQKQIAGL--TPDLAP-----LOIAQOLQNDIDOPTSTGMTSAS 1275  
QY 682 CASGNTQILCISTK-----NLNTQAMVGGTGLVGGIYEDNGNTLTKVPL 728  
Db 1276 IAAFNEKLSAARTKQIEDRLVLAHPDVATIRQNVTAANAASALDQARNGLTVDKAPL 1334

RESULT 8

US-09-815-242-12610  
; Sequence 12610, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlson, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA, Olla  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12610  
; LENGTH: 5795  
; TYPE: PRF  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12610

Query Match 4.3%; Score 166.5; DB 9; Length 5795;  
Best Local Similarity 19.9%; Pred. No. 0.0045;  
Matches 175; Conservative 117; Mismatches 324; Indels 265; Gaps 37;

QY 34 SLPNKQKIV-----KVSFDEIYNPTGFTSSPARIALDFEQTGISMDOQVLEYADPL 86  
Db 2387 TIANKPDYVTLDAHTGKVTFNANTVKPSAITITPK-----AGRG----- 2426  
QY 87 LSKISAAQNSPARLVNLNPKFQYNTVEVRGNKWMIFINESDDTVSAPARPAVKAAP 146  
Db 2427 ---HSASSNPS-----TLTAPATHV---NTTEIVKDGNSVNTAAEINNVAQVA----- 2469  
QY 147 AKQOCRTYVQVSIRIQLYLGKTTAAAPFT-----ESVVSVAAPSPAKQ 194  
Db 2470 ---NKRATIKNGTAMPINLAGGSTTIPVTYVNDGSTVEQESI-----FTKADKR 2519  
QY 195 AAASAKQOJTAAP-----AKQO---TAAPAKQQA-----AAPAKQTN 227  
Db 2520 ELITAKNHLDDPVSTDGKKPGTITQYNNAIHNAQQOINTAKTEAQOVINNERATPQQVSD 2579  
QY 228 IDFRKDGKNAGIIEALAGFAGQPDISQOHDHIVTLKN-----HTLPFT---LQSLD 278  
Db 2580 ALTKVRAAQTKINEAKAL-LQNKENSQ-----LVTSKNLIQSSVYNQVPSTTGMTQOSID 2633

QY 279 VADPK-----TPQKV-----TLKRLNNDTQLIITAGNWELVNK-SAAPGYFTFQ 323  
Db 2634 NYNAKREARTEITAAQRVIDNGDATAQQISDEN-----TTAQRNDLTNQISQATNLAAVE 2689  
QY 324 VLPKKQNLSEGGVNN---APKFTTGRKISLDFQDVE-----IRTIQLILAKESGM 370  
Db 2690 SVKQSANSLDGMGNLQTAINDKSGTLASQNFLDADBEQKNAYNQAVSNAETILNKQTGP 2749  
QY 371 NIVASDSVNGKMTLSLKDVPDQALDLYMCAW-----NLDMRQOQGNIVNIAPIRELLA 423  
Db 2750 N-TAKTAV-----EOALNNVNSAKGALNGTQNLNNAKQAAITAINGASDLNQ 2795  
QY 424 KDKAFQAEMDIADLGALYSQNFQKXVVEEPRSIL-----RLDNAD 466  
Db 2796 KQDALKAAQAN---GAQRVSNAAQDVQRNATELNTAMGLTKHAIADKTNLASSKYVNAD 2851  
QY 467 TT-----GHRNTLVSGRGSVLDPATNTLIVTDTSRVIEKFR----- 503  
Db 2852 STKQNAVTTKVTNAEHIIISGTPVTTPSEVTAANQVNSAKOELNGDERLRAVKQNAVNT 2911  
QY 504 --KLIDELDVPAAQVMEARIVEA-----ADGFSRDLGVKFG 538  
Db 2912 AIDALTQLTNP-QKAKLKEQVGQANRLIEDIQTVTNGQALNNAMKGLRDSIANETTVKAS 2970  
QY 539 ATGKKKLKNDTSAPFGWGVNSGFGDDKWAETKINLPITAAANSISLVRAISSGALNLEL 598  
Db 2971 QNYTDSAPNNQSTYNSAVSNAKGIINQTNPTMDTSAITQATTVNNAKKGLNGAENLRN 3030  
QY 599 SASLSLKTXTLANPRLVLTORKEA---KTESGYEIPFTVTSIANGSGSNTTELKXAVLG 655  
Db 3031 AQNTAKQNLTLSH---LTNNQKSAISSQIDRAGHVS-EVTAAKNAATELNTONGNLEQA 3086  
QY 656 LTVPTNPITPDGQIIMTVKINKDSPAQACAGNQTLICISTKTLNT-----QAMVENGOTLI 710  
Db 3087 IHDQNTVKQGVNFTDADKAKRDAYTNVNSRAETIL-NKTQGANTSQDVEEAQN----- 3140  
QY 711 VGGIYEDNGNTLTKVPLGDIPIVGNLTKTRCKTKDRREL 751  
Db 3141 -----VTSAKNALNGQNV-----TNAKTAKUAL 3165

RESULT 9

US-10-193-764-45  
; Sequence 45, Application US/10193764  
; Publication No. US20030133943A1  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH  
; FILE REFERENCE: 1038-1239MIS  
; CURRENT APPLICATION NUMBER: US/10/193,764  
; CURRENT FILING DATE: 2002-07-12  
; PRIOR APPLICATION NUMBER: 09/167,568  
; PRIOR FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 91  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 45  
; LENGTH: 1073  
; TYPE: PRF  
; ORGANISM: Haemophilus influenzae  
US-10-193-764-45

Query Match 4.2%; Score 160.5; DB 12; Length 1073;  
Best Local Similarity 19.7%; Pred. No. 0.00031;  
Matches 125; Conservative 88; Mismatches 212; Indels 211; Gaps 25;  
QY 205 APAKQOCTAAPAKQOAAAPAKQTNIDFRKDGKNAGIIEALAGFAGQPDISQOHDHIVTL 264  
Db 8 APSAERT-----DTGEDVEYTGTDADI-----NHQKQNSKSTL 42





```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12996
; LENGTH: 6281
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12996
```

```
Query Match
Best Local Similarity 4.1%; Score 159.5; DB 9; Length 6281;
Matches 169; Conservative 100; Mismatches 327; Indels 223; Gaps 38;

QY 23 SAGNITDIKVSLLPNKQKIVKVSFQKE-----IVNPTGFVSSPARIALDFEQ 70
DB 2549 STAKDDVRAK-----QPIVDADRKQNAYNATVNAENINATSOPTLDFSAVQAANQ 2603
QY 71 TGISMDOQVLEVADPLLSKISAQNSSRARLVNLNPKQVYNEVRG-----NKKWIF 123
DB 2604 --VSTNKTALNGAQNLANKKQETTANINQSLNNAQAQDLNTQVTNAPNISTVNOVKTK 2661
QY 124 INESDQTVSA-----PAPPAKAPAAAPAKQOQGRVYQVRSIRIQLYFGKTTAA 174
DB 2662 AQLOQAMERLINGQDKQVQKVSQVNTDADPEK-----QTAYNAVNTAA 2706
QY 175 APFTSVSVSAPFSPAKQAAASAKQQTAPAKQQTAPAKQQAAPAKQTNIDFRKDG 234
DB 2707 ---ENIIN-----QANGTNANQSQVEAALSTVTTTKQALNGDRKVTD---AKN 2749
QY 235 KNAGIIEAALGFAQO---PDISQCHDHIIVTLKNHTLPTTLQRSLOVADFKTPVQKVT 290
DB 2750 ANQTLSTDLNNAQKQAVTGMINQAH-----TVAEVTOAIQ--T 2787
QY 291 LKRLNNDTQLIITTAGNWE--LVNKAAPGYFTF-QVLPKQKNSGGVNNAPKFTTGRK 347
DB 2788 AQELN-----TAMGNLNSLNDKDTLLGSQNFADADPEKKNAYNEAVHNAE----- 2833
QY 348 ISLDQDVVEIRILQILAKESGMNI-----VASDSVNG-KWTLKQVDPWDQALDLYMQ 400
DB 2834 -----NILNKSTGTVNPKQOQVEAMQVNAATKAALN-----QTQNLKAKQ 2874
QY 401 ARNLMROQGNIVNIAPRDELLAKDKAFLEAKEDLADLGALYSQNFQKYNVEFRSIL 460
DB 2875 HANTAIDGLSHLTN-AQKEAL---KQLVQOQSTTVAE-----AQNEQKANNVDAANDKL 2924
QY 461 R---LDNADTGNRNTLVSGRSVLIDPATNLTIVDTFRSIVIEKPRKILDELDPVPAQVM 517
DB 2925 ROSIADNATTKQNYTDASQNK---KDAYNNAVTT-----AQGIIDQTTSTPT----- 2969
QY 518 IEARIVAAADGFSRDLGVKFGATGKKLK--NDTSAGFGVNSGFGGDDKNGAETKINLP 575
DB 2970 LDPTVINQAG--QVSTTKVNLNGNENLEAAKQQAQSQSLGSLDLNNAQKQTVTDQIN-- 3025
QY 576 ITAANSISLVRAISGALNLELS-----ABESLSKTKTLANPRVLTONRKEAKIESGY 629
DB 3026 ---GAHTVDEANQIKQNAQNLNTAMGNLQAIADKATKATVNFDDAQAKQA-----Y 3077
QY 630 EIPFT----VTSIANGSGSTNTELKAVLGLTVTPNITPDGQIMTVKINKDSPAQACASG 685
DB 3078 NTAVTNAENIISKANGGNATQAEVQAQKQVNAQ-----ALNGNAVVOHAKD 3126
QY 686 NOTILCISTKNLN-----TQAVWNGGTLIVGGI-----YEDNGNTLTKVPLLG-- 730
DB 3127 EATALINSSNDLNAQKQALQKQVQNAIT--VAGVNNVQKQTAQELNNAMTQKOGIADKE 3184
QY 731 DIPVTGNLFKRGKTD-----BRELLIFITPRIMGT 762
DB 3185 QTKADGNFVNAQDPDKQNAVNAQVAKAELISATPDVVVT 3223
```

```
RESULT 12
US-09-815-242-5883
; Sequence 5883, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5883
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5883
```

```
Query Match
Best Local Similarity 4.1%; Score 159; DB 9; Length 837;
Matches 175; Conservative 129; Mismatches 303; Indels 308; Gaps 42;

QY 1 MNTKLTIIISGLFVATAAFOTASAGNITDIKVSLLPNKQKIV--KVSFDEI VNPFGFVT 58
DB 25 LNTAMTALKEA--IADKA-ETKASGNY---VNADANKQAYDEKVTAAENIUSGTPTPT 77
QY 59 SSPARIALDFEQTGISMDOQVLEVADPLLSKISAQNSSRARLVNLNPKQVYNEVRGN 118
DB 78 LTPADVT-----NAAQTQVNTAKTQLN---GNHNLVAKQ 108
QY 119 KWIIFNESDD---TVSAPARPAKAPAKAQOQ-CRTVYQVRSIR-----IOTLYPG 169
DB 109 NA-----NTAIDGLTSLNGPQAKLK-----EQVQATTLPNVQTVRDNAQTINTAMKG 157
QY 170 KTTAAAPFTESVSVSAPFSPAKQ-----QAAASAKQ---QTAAPA-KQQTAAAPAKQ 218
DB 158 LRDSIA--NEATIKAGQNYTDASQNKQTDYNSAVTAAKAIIGOTTSPSMNAQEINQAKDQ 215
QY 219 AAAPAKQTNIDFRKDKGNAGIIEAALGFAGQPDISQCHDHIIVTLKNHTLPTTLQRSLD 278
DB 216 VTAKQOALN-----GQENL-----RTAQTNAKOHLN 241
QY 279 VADFKTPVQKVTLLK-----LNNDTQLIITTAGNWLNVKSAAPGYFTFVLPKQNL 331
DB 242 GLSLDITDAQKDAVKRQIEGATHVNEVTQ---AQNNADALNTAMT-----NL 284
QY 332 ESG-----GVNNAPKFTTGRKISLDFQVDE-----IRTILOILAKESGMNI 372
DB 285 KNGIQDQNTIKQGVN-----FTDADEAKRNAYTNVATQAEQILNKAQGN- 329
QY 373 VASDSVNGKMTLSLKDVPMQDQALDLYNQARNLDMRQOQGNIVNIAPRDELLAKDKAFQAE 432
DB 330 TSKDGV-----ETALENVQRAKN-ELNGNQNVAN-----AKTTAK 363
QY 433 KDIADLGALYSQNFQKYNVEFRSILRLDNADTTGNR-NTLVSGRGSVLIDPATNTLI 491
```

Db 364 NALNNLTSSINNAQKALKSQIEGATTVAGVNVQVTTASELNTAMSNLQNGINDEAATKAA 423  
 QY 492 VTDRSV-----IEKFKLIDELDVPAQVQVMEARIVEAADGFSRDL----- 533  
 Db 424 LNTAMGNLQKAIADKATKATVNFETDADQAK-QQAYNTAVTNAENIISKANGGNATQAEV 483  
 QY 534 ----GVKFGATGKKK-----LKNDSAFGWGVNSGFGDDKQWGAETKINLPITAAANSI 583  
 Db 484 KLRQSIADNATTKQNVYTDASQNKQDAYNNAVTTAQGIIDOTTSPILDPVINAAGQV 543  
 QY 584 SLVRAISSGALNLELS-----ASELSKTKTLANPRVLT-----QNRK- 621  
 Db 544 STTKNALNGNENLEAAKQQAQSSQSLDNLNNAQKQVTDQINGAHTVDEANQIKQNAQN 603  
 QY 622 -----EAKIESGYEIPFT-----VTSIANGSGSSTNTEL 649  
 Db 604 LNTAMGNLQKAIADKATKATVNFETDADQAK-QQAYNTAVTNAENIISKANGGNATQAEV 662  
 QY 650 KXAVLGLTVTPNITPDGQIIMTVKINKDSPACQASGNOTILCISTKNLN-----TQAMV 703  
 Db 663 EQAIKQVNAKQ-----ALNGVANVQHAKDEATALINSNDLNQAKDALKQOV 711  
 QY 704 ENGOTLIVGGI-----YEDNGNLTQKVPILG-DIPVIGNLFKTRGKKT----- 747  
 Db 712 QNATT--VAGVNVVQKQTAQELNNAMTOLKQGIADKEQTKADGNEFVNADPDQONAYNOAVA 769  
 QY 748 RRELLIFITPRIMGT 762  
 Db 770 KAEALISATPDVVVT 784

RESULT 13

US-09-815-242-13080  
 ; Sequence 13080, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; TITLE OF INVENTION: Prokaryotes  
 ; FILE REFERENCE: ELITRA.011A  
 ; CURRENT APPLICATION NUMBER: US/09/815,242  
 ; CURRENT FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: Fast-Seq for Windows Version 4.0  
 ; SEQ ID NO 13080  
 ; LENGTH: 875  
 ; TYPE: PRN  
 ; ORGANISM: Staphylococcus aureus  
 ; US-09-815-242-13080

Query Match 4.1%; Score 159; DB 9; Length 875;  
 Best Local Similarity 19.1%; Pred. No. 0.00087;

Matches 175; Conservative 129; Mismatches 303; Indels 308; Gaps 42;  
 QY 1 MNTKLTIKISGLFVATAAFQTASAGNITDIKYSSLPNKQKIV--KUSFKKEIYNVPGFVT 58  
 Db 62 LNTAMTALKRA--IAKA-ETKASGNY--VNADANKQAVDEKVTAAENIVSGTPTPT 114  
 QY 59 SSPARIALDFEQTGISMDQOVLEYADPILLSKISAAQSSRRARLVNLNKPQGVNTVEVRGN 118  
 Db 115 LTPADV-----NAATQVNTAKTQLN-----GNHLEVAQK 145  
 QY 119 KWIFINESDD---TVSAPARPAVKAAPAKAQOG- CRTVYQVRSIR-----IQTLYPG 169  
 Db 146 NA-----NTAIDGLTSLNGPQKAKLK-----EQVQATTLFNVQTVRDNQATLNTAMKG 194  
 QY 170 KITTAAPFTTESVVSVAPEPAKQ-----QAAASAKQ-----QTAAPA-KQOTAAAPAKQ 218  
 Db 195 LRDSIA--NEATIKACQNVYTDASQNKQTDYNSAVTAAKAIIGOTTSPSNMAGNEINQAKQ 252  
 QY 219 AAPAKQTVNIDFRKQGNAGIIEAALGAFAGQPDISQQHDHIIIVTKNHLPTTLQSRSD 278  
 Db 253 VTAKQQAALN-----GOENL-----RTAQTNKQHLN 278  
 QY 279 VADFKTPVQKVTLKR-----LNNDTQLIITTAGNWEVLVNSKAAPGYFTFQVLPKQNL 331  
 Db 279 GLSDLTDAQKDAVKQIEGATHVNEVQ-----AQNADALNTAMT-----NL 321  
 QY 332 ESG-----GVNNAPKFTTGRKISLDFQV-----RTILQILAKESGMNI 372  
 Db 322 KNGIQDONTIKQGVN-----FTDADEAKRNAYTNAVTTQAEQILNKAGQPN- 366  
 QY 373 VASDSVNGKMTLSLKDVPMDQALDLVMQARNLDMRQCGNIVNIAPRDELLAKDKAFLOAE 432  
 Db 367 TSKDGV-----ETALENVQRAKN-ELNGNQNVAN-----AKTTAK 400  
 QY 433 KDIADIGALYSQNFQKYKNVEEPRSIILRLDNADTTGNR-NTLVSGRGSVLDPATNTLI 491  
 Db 401 NALNNLTSSINNAQKALKSQIEGATTVAGVNVQVTTASELNTAMSNLQNGINDEAATKAA 460  
 QY 492 VTDRSV-----IEKFKLIDELDVPAQVQVMEARIVEAADGFSRDL----- 533  
 Db 461 LNTONLEKAKOHANTAIDGLSHLTNAQKALKQVQOSTTVAEAGQNEQKANNYDAAMD 520  
 QY 534 ----GVKFGATGKKK-----LKNDSAFGWGVNSGFGDDKQWGAETKINLPITAAANSI 583  
 Db 521 KLRQSIADNATTKQNVYTDASQNKQDAYNNAVTTAQGIIDOTTSPILDPVINAAGQV 580  
 QY 584 SLVRAISSGALNLELS-----ASELSKTKTLANPRVLT-----QNRK- 621  
 Db 581 STTKNALNGNENLEAAKQQAQSSQSLDNLNNAQKQVTDQINGAHTVDEANQIKQNAQN 640  
 QY 622 -----EAKIESGYEIPFT-----VTSIANGSGSSTNTEL 649  
 Db 641 LNTAMGNLQKAIADKATKATVNFETDADQAK-QQAYNTAVTNAENIISKANGGNATQAEV 699  
 QY 650 KXAVLGLTVTPNITPDGQIIMTVKINKDSPACQASGNOTILCISTKNLN-----TQAMV 703  
 Db 700 EQAIKQVNAKQ-----ALNGVANVQHAKDEATALINSNDLNQAKDALKQOV 748  
 QY 704 ENGOTLIVGGI-----YEDNGNLTQKVPILG-DIPVIGNLFKTRGKKT----- 747  
 Db 749 QNATT--VAGVNVVQKQTAQELNNAMTOLKQGIADKEQTKADGNEFVNADPDQONAYNOAVA 806  
 QY 748 RRELLIFITPRIMGT 762  
 Db 807 KAEALISATPDVVVT 821

RESULT 14

US-09-815-242-5835  
 ; Sequence 5835, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert

```

; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5835
; LENGTH: 2434
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5835

Query Match 4.1%; Score 158; DB 9; Length 2434;
Best Local Similarity 20.5%; Pred. No. 0.0054;
Matches 158; Conservative 99; Mismatches 303; Indels 212; Gaps 36;

QY 23 SAGNITDIKVSRLPKQKIVKVSFDKE-----IVNPTGFTVSSPARIALDFEQ 70
DB 1770 STANKDDVKAS-----QPYVDADRDKQNAVYTAVTNAENIINATSQPTLDPSTQAQ 1824

QY 71 TCISMDQOVLEADPLLSKISAAQNSSRARLVNLNPKGQYNTVRG-----NKYWF 123
DB 1825 --VSTKTLNGAQLANKKETTANINQLSHLNNAQKODLNTQVTHAPNISTVNQVTK 1882

QY 124 INESDDTVSA-----PARPAVKAAAPAAKQOQCRVYQVRSIRIQLYPGKTTAA 174
DB 1883 AEQLDQAMERLINGIQDKQVQKQSVNFTDADPEK-----QYAVYNAVTA 1927

QY 175 APFTESVVSVAPEKQQAASAKQOATAAPAKQOAAAPAKQOTNIDFRKDG 234
DB 1928 ----ENIIN-----QANGNANOSQVEAALSTVTTTKQALNGDRKVD---AKNN 1970

QY 235 KNAGHIELAALGFAQ----PDISQOHDHIIIVTLKNHTLPTTLQSLDVADEFKTPQKVT 290
DB 1971 ANQTLSTDLNNAQKQAVTGNINQAH-----TVAEVTQAIQ--T 2008

QY 291 LKRLNNDQLIITAGNWE--LVNKSAPGYFTF-QVLPKKNLESQGVNNAKPTFTGRK 347
DB 2009 AQELN-----TANGNLKNSLNDKDTTLGSCNQNFADADPEKKNAYNEAVHNAE----- 2054

QY 348 ISLDPQDVEIRILQILAKESGMNI-----VASDSVNG-KWTLSLKDVDPWDQALDVMQ 400
DB 2055 -----NILNKSTGTNVPKDQVEAAVQVNAATKAALN-----GTQNLERAKQ 2095

QY 401 ARNLDMROCGNIVNIAPDELLAKDAFLQAEKIDIALGALYSQFQLKYNVVEFRSIL 460
DB 2086 HANTAIDGLSLHTN-AQKEAL-----KQLVQOSTTVAE-----AQGNQKANNVDAAMDKL 2145

QY 461 R---LDNADTTCNRNTLVSGRGSVLIDPATNTLIVTDRSVIERFKRLIDELDVPAQVM 517
DB 2146 RQSIADNATTQONQNYTDASQNK---KDAYNAVTT-----AQGIIDTTSPT----- 2190

```

```

QY 518 IEARIVEAADGFSRDLGVKFGATGKKLK--NDTSAFGMGVNSGFGDDDKWAETKINLP 575
DB 2191 LDPTVINQAG--QVSTTKNALNGNENLEAAKQOASQSLGDLNNAQKQVTDQIN-- 2246

QY 576 ITAAANSISLVRASISGALNLELS-----ASELSKTKTLANPRVLITONRKEAKIESCY 629
DB 2247 ---GAHTVDEANOIKONAQNLTAMGNLKQAIADKDKATKATVNFDDADQKQQA-----Y 2298

QY 630 EIPFT----VTSIANGSGSSTTELKXAVLGLVTPTNITPDGQIIMTVKINKDSPAQCSG 685
DB 2299 NTAVTNAENIISKANGGNATQAEVQAIKQVNAAQ-----ALNGNANVQCHAKD 2347

QY 686 NQTILCISTKNLN-----TQMVENGTLIVGGIYE-----EDNGNTLTKV 726
DB 2348 EATALINSSNDLNQAKDALKQOVQNAFT--VAGVNVNKQTACELNNAQTOL 2397

RESULT 15
US-10-193-764-65
; Sequence 65, Application US/10193764
; Publication No. US20030133943A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 65
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-193-764-65

Query Match 4.0%; Score 155.5; DB 12; Length 1095;
Best Local Similarity 19.1%; Pred. No. 0.0024;
Matches 146; Conservative 123; Mismatches 284; Indels 213; Gaps 41;

QY 25 GNITDIKVSRLPKQKIVKVSFDKEIVNPTGFTVSSPARIALDFEQTIGISMDQVLEYAD 84
DB 126 GNI-----NITAKODIAFEKGSNOVITGQITISG-----NQKGFERN----- 164

QY 85 PULLSKISAQNSRRARLVNLNPKG---QYNTFVRGNKVIWIFINESDDTVSAPAPAKVA 141
DB 165 -----VSLNGTSGGLQFTTK-RTNK-YAITNKPEGLNTISGK--VNI 202

QY 142 APAAPAKQOQC-----RTVYQVRSIRIQLYPGKTTAAAPFTESVVSVAPEKQQA 196
DB 203 SMVLPKNESGYDKFKGRTYWNLTSLNVSSEGFNLIDSRGSDSAGTLTQPIN----- 255

QY 197 ASAKQOATAAPAKQOATAAPAKQOTNIDFRKDGKNAGIIEALAALGFA----- 248
DB 256 -----LNGISFNKDTTFNVVERNA-----RVNFDIKAP-----IGINKYSSLNYSFNGNISV 302

QY 249 ---GQPDISQOHDHIIIVTLKNHTLPTTLQSLDVADEFKTPQKVTKLNNNDT 298
DB 303 SGGGSVDFT-----LLASSNVQTPGVNLSKYFNVSTGSSLRFKTSKTKGFSIEKDL 357

QY 299 QLIITTAGNWLWNKSAAPGYFTFQVLPKKQ-NLESQGVNNAKPTFTGRKISLDFO-DVE 356
DB 358 TL-NATGGNITLLQVEGTGDMIGKIVAKKNITFEGGNI-----TFGSEKAVTEIEGNYT 411

QY 357 IRTILQILAKESGMNIVASDSVNGKVTLSLKDVPWDQALDVAQENLDMROCGNIVNIA 416
DB 412 IN-----NNANVTILIGSDFDNHQPPLTIKK-----DVIINSGNL--TAGGNIVNIA 455

```

```
QY 417 PRODELLAKDAFLQAEKDIA-DLGALYSQNFOLKYKVEEFSILRLDNADTTGNRTLV 475
Db 456 --GNLTVESNANFKAITNFTNVGGLF-----DNKGSNLSI 490
QY 476 SGRGSVL--IDPATNTLIVDTRSVIEKFRKLID-----ELDVPAQ 514
Db 491 AKGGARFKDIDNSKLSITNNSST--YRTIISGNITNKGDNLNITNEGSDTEWQIGD 547
QY 515 QVMIEARIVEAAD--GFSRDGLVKFGATGKKLKNDTSAFGWGVNSFGGDDKWAETKI 572
Db 548 VSQKEGNLTSSDKINIKTKQITKAGVDGENSDATN----- 585
QY 573 NLPITAAANSISLVRAS-SGALNLELSASE---SLSKTKTL--ANPRVLTONR-KEAK 624
Db 586 NANLTIKTKELKLTQDLNLSGFNKAETAKGSDLTIGTNSADGTNAKVTNFQVKDSK 645
QY 625 IES-GYEIPFTVTSIANGSGSTNTE-LKKAVALGLTV-TPNITPDGQIIMTVKINKDSPAQ 681
Db 646 ISADGHKVTLLH-SKVETSGSNNTEDSSDNNAGLTIDAKNVTNNITS----HKAVSIS 700
QY 682 CASGNQTLICISTKNL---NTQAMVENGTLIVGGIYEEDNGNTLT 724
Db 701 ATSGEITTKGTITNATGNVEITATQGS--ILGGIESSSGSVTLT 744

RESULT 16
US-10-193-764-63
; Sequence 63, Application US/10193764
; Publication No. US20030133943A1
; GENERAL INFORMATION:
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 1536
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-193-764-63

Query Match 4.0%; Score 155.5; DB 12; Length 1536;
Best Local Similarity 19.1%; Pred. No. 0.0041;
Matches 146; Conservative 123; Mismatches 284; Indels 213; Gaps 41;

QY 25 GNITDIKVSSLPNKQIKVKYSPDKIEIVNPTGFTVTSSPARIALDFEQTGISMDQOQVLEYAD 84
Db 567 GNI-----NITAKQDIAFEKGSNQVITGQGTITSG-----NKGFRFNN----- 605
QY 85 PLLSKISAAQNSSRARLVNLNKKPG---QYNTVEVRGNKVKWIFINESDDTVSAPAPPAVKA 141
Db 606 -----VSLNGTSGGLQFTTK-RTNK-YAINTKFEGLTNISGK--VNI 643
QY 142 APAAPAKQOQC-----RTVYQVRSIRIOTLYPGKTTAAAPFTESVVSAPPSPAKQAAA 196
Db 644 SMVLPRNESGYDKPKGRTYWNLTSLNVSSEGEFNLTDISRGSDSAGTLTQPVN----- 696
QY 197 ASARQQAATAAPAKQAAAAPAKQTNIDPRKDGKNAGIIEAALGFA----- 248
Db 697 -----LNGISFNKDTTNVERNA-----RVNFDIKAP---IGINKYSSLNAYSFNGNISV 743
QY 249 ---GQPDISQOHDHIIIVLKNHT-PTTLQRS--LDVAD-----FKTPVQKVTLKRLNNDT 298
Db 744 SGGGSVDFT-----LLASSSNVQTPGVVINSKYFNVTSGSLRFTKTSKGTGFSIEKDL 798
QY 299 QLIITTAGNWLNVKSAAPGVFTFQVLPKKQ-NLESGGVNNAKPTFTGRKISLDFQ-DVE 356
```

```
Db 799 TL-NATGNITLLQVEGTDGMIGKIVAKKNITFFGGNI-----TFGSKAVTEEGNVT 852
QY 357 IRTILQILAKESGMNIVASDSVNGKMT-SLKDVDPWDQALDLVMQARNLDMROQGNIVNIA 416
Db 853 IN-----NNANVTLIGSDFNHQKPLTIKK-----DVIINSGNL--TAGGNIVNIA 896
QY 417 PRODELLAKDAFLQAEKDIA-DLGALYSQNFOLKYKVEEFSILRLDNADTTGNRTLV 475
Db 897 --GNLTVESNANFKAITNFTNVGGLF-----DNKGSNLSI 931
QY 476 SGRGSVL--IDPATNTLIVDTRSVIEKFRKLID-----ELDVPAQ 514
Db 932 AKGGARFKDIDNSKLSITNNSST--YRTIISGNITNKGDNLNITNEGSDTEWQIGD 988
QY 515 QVMIEARIVEAAD--GFSRDGLVKFGATGKKLKNDTSAFGWGVNSFGGDDKWAETKI 572
Db 989 VSQKEGNLTSSDKINIKTKQITKAGVDGENSDATN----- 1026
QY 573 NLPITAAANSISLVRAS-SGALNLELSASE---SLSKTKTL--ANPRVLTONR-KEAK 624
Db 1027 NANLTIKTKELKLTQDLNLSGFNKAETAKGSDLTIGTNSADGTNAKVTNFQVKDSK 1086
QY 625 IES-GYEIPFTVTSIANGSGSTNTE-LKKAVALGLTV-TPNITPDGQIIMTVKINKDSPAQ 681
Db 1087 ISADGHKVTLLH-SKVETSGSNNTEDSSDNNAGLTIDAKNVTNNITS----HKAVSIS 1141
QY 682 CASGNQTLICISTKNL---NTQAMVENGTLIVGGIYEEDNGNTLT 724
Db 1142 ATSGEITTKGTITNATGNVEITATQGS--ILGGIESSSGSVTLT 1185

RESULT 17
US-10-092-880-2
; Sequence 2, Application US/10092880
; Publication No. US20020164354A1
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
; TITLE OF INVENTION: HAEMOPHILUS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/092,880
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: 09/155,614
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 08/617,697
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: PCT/US97/04707
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1536
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-092-880-2

Query Match 4.0%; Score 155.5; DB 14; Length 1536;
Best Local Similarity 19.1%; Pred. No. 0.0041;
Matches 146; Conservative 123; Mismatches 284; Indels 213; Gaps 41;

QY 25 GNITDIKVSSLPNKQIKVKYSPDKIEIVNPTGFTVTSSPARIALDFEQTGISMDQOQVLEYAD 84
Db 567 GNI-----NITAKQDIAFEKGSNQVITGQGTITSG-----NKGFRFNN----- 605
QY 85 PLLSKISAAQNSSRARLVNLNKKPG---QYNTVEVRGNKVKWIFINESDDTVSAPAPPAVKA 141
Db 606 -----VSLNGTSGGLQFTTK-RTNK-YAINTKFEGLTNISGK--VNI 643
QY 142 APAAPAKQOQC-----RTVYQVRSIRIOTLYPGKTTAAAPFTESVVSAPPSPAKQAAA 196
Db 644 SMVLPRNESGYDKPKGRTYWNLTSLNVSSEGEFNLTDISRGSDSAGTLTQPVN----- 696
```



```
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Fusobacterium necrophorum
; US-09-841-786-4

Query Match
Best Local Similarity 3.9%; Score 150.5; DB 9; Length 714;
Matches 143; Conservative 94; Mismatches 267; Indels 173; Gaps 32;

QY 121 WIFINESDDTVSAPARPA-----VKAAPAAPAKQOQCRTVYQVRSIRIOTLYPGKTTAAA 175
Db 54 WNHGSGNSNISAGLAGAAVANNIQKTSALVKNUSDIR--NANKFKNALSGGTQVAAG 110
QY 176 PFTESV-----VVSAPFSAKQQAASAKQQTAAAPAKQQTAAAPAKQQAAPAK 224
Db 111 AGLEAVKESGGQKSVLLGTSASINLVNVEVSAKSENNTVA-----GESESQ 157
QY 225 QTNID---FRKDGKNAGIIEAALGALFAGOPDISQQDHHIIVTLKXHTLPTTLQRSILDVAD 281
Db 158 KMDVDVTAQADQVITGALNLQ---AGK-----SNGTVGAT----- 190
QY 282 FKTPVOKVTLKRLNNDTQLIITAGHWELVYKSAARGYP-TFOV-----LPKQNLSESG-QV 336
Db 191 -----VTVAKLNKNVNASI-SGGRYTNVNRADAKALLATTQVTAATVTTGGTISGAGL 242
QY 337 NNAPKFTGKRSILDFQDVEIRTILOILAKESGMNIVASDSVNGKMTLSLKDVPMDQALD 396
Db 243 GNYQGAVSYNKID--NDEASVDKSSIEGANEINVIAXD-VKGSSDLAKE---YQAL- 293
QY 397 LVNQARNLDMRQOQNIUNTAAPRELLAKDKAFLOAEKDIADLIGALYSQNFQLYKRVNEEF 456
Db 294 -----LNGKOKKYLE-DRGINTTNGYVYTKEQLEKAKKKKG 328
QY 457 RSILRLDNADTTGNRNLTLYSG-----RGSVLIDPATNTLIVTDRSVEIKFRKLIIDELVDP 512
Db 329 AVIV-----NAALSVAAGTKSAGGVAI--AVNT-VKNKFKALSOSNKEAGEDKIH 376
QY 513 AQOVMIEAR-----IVBAAG--FSRDLGVKFGATGKKKLKNDTSAPFGWVNSGFGGDDKW 566
Db 377 AKHVNVEAKSSTVYVNAASGLAISKDAFSGMGSGAWQDLSNDTIA---KVDKGRISADSL 433
QY 567 GAETKINLPITAAANSI--SLVRAISSGALNLSASLSKTKTLANPRVLTQNRKEAK 624
Db 434 NVNANNSILGVNAGTIAGSLSTAVGAAAFANNVTLHNKTSALITGKNP-----FSGKNTK 489
QY 625 IESGYEIPFTVTSIANGSGSTNTTELKAVLGLTVTPN-----ITPDGQIMTVKINK 676
Db 490 VNVQALNDSHITNVSAGGAAS---IKQAGIGGMVSVNRGSDTEALVSDSEFEGVSSFNV 546
QY 677 DSPAQCASGNTILCISTKNLNTQAMVENGGLTVGGIYEEDNGNTL--TKVPLLLGDIPV 734
Db 547 DAKDQ-----KTINTIANGGK--AAGV-----GATVAHTNIGKQSVIAI 586
QY 735 IGNLFKTRGKKTDRREL 751
Db 587 VKNSKITTANDQDRKNI 603

RESULT 20
US-09-815-242-5635
; Sequence 5635, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Olesen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
```

Db 1229 -----NITATD-----DNGVDTAKDAGKNSIQSTQAPATA-----VKSNAKND 1266  
QY 572 INLPITAAANSISLVRAISSGALNLSASESL-SKTKTLANPRVL-----TONRKE 622  
Db 1267 VQAVTTQNAID-----NITGATTEERNAKDLVLKAKKAYODILNAQTNDVTQIKQ 1322  
QY 623 AKIE-SGYEIPFTVTSIANGSSSTNTELKXAVLGLVTPNITPDGQIIMTVKINKDSPAQ 681  
Db 1323 AVADVQGITADTTIKDVAKDELATKAREQKALIAQTA-----DATTEEKEQANQOVDAE 1376  
QY 682 CASGNQTLICISTKMLNTQAMVENGGLIVGGIYEDNGNLTQKVPLLGDIPIVGNLFT 741  
Db 1377 LTQGNQNI-----ENASQI-----DDVNTAKDNAIQADIPQAS---T 1411  
QY 742 RGKKTDRRELLFITPRIMGTAGNS 766  
Db 1412 DVKTNAAREALLTEMQNKITEILNNN 1436

RESULT 21

US-09-815-242-12389  
; Sequence 12389, Application US/09815242  
; Patent No. US2002061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlson, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815.242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12389  
; LENGTH: 2368  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12389

Query Match 3.9%; Score 150.5; DB 9; Length 2368;  
Best Local Similarity 19.2%; Pred. No. 0.021;  
Matches 166; Conservative 119; Mismatches 327; Indels 253; Gaps 36;  
QY 6 TKISGLFVATA-----AFQTSAGNI-----TDIKUSSLPNKQIKVVSFQXEIVNP- 53  
Db 721 TRAKIDDAATNAQVEAIKTKAINDINQTPATTAKAAALEEFVEVVQAIQDAPLND 780  
QY 54 -----TGFTSSPARIALDFEQT---GISMDQQVQVLEYADPLLSKISA 92  
Db 781 TTNEEVAERIERINAAGVGVKATEATTAGDLERVKNEEISKIENITDSTQTKMDAYNE 840  
QY 93 AQNSSRRLVLNLKPGQYNEVRGNKVNIFINESDDTVTSAPARPAKAAAPAKQGC 152

Db 841 VKQATARKTQNTATVSNATNEEV-----AEADAAVEA-----AQKQGL 878  
QY 153 RTVQVRSIRIQTLYPGKTTAAAPFTESVSVSAPFSPAKQOAAASAKQOATAAPAKQ-- 210  
Db 879 HDIQVKS-----KQEVADTKSVLDKINALIQOAKVKPAADTEVENAYNTRKQEI 929  
QY 211 ---TAAPAKQOAAAPAKQNTIDFRK-----DGNAGIETELAALGFAGQPDISOQHDHI 260  
Db 930 QNSNASTTEEEKQAA---YTELDTKKQEARTNLDAANT-----NSDVTITAKONG 974  
QY 261 IVTLKNHSLPTTLORSILDVADF--KTPVQKVTLKRLNNDT-----QLIITTAG 306  
Db 975 IAAI-NQVQAATTKSDAKAEIAQKASERKTAIEMNDSTTEEQOAAKDKVDQAVVT--A 1031  
QY 307 NWELVNSAAPGYFTFQVLPKKQNLSEGGVNNAPKFTTGRKISLIDFQVEIR-TILQILA 365  
Db 1032 NADIDNAAA-----NTDVEDNA-KTTNEATIAAATTPDANVKYAKQAIA 1073  
QY 366 KESGMNIVASDSVNG-----KMTLSLKDVDPWDOALDLVMOARNLDMRQOQNI-VNI- 415  
Db 1074 DKVQAQETAIIDANGATTETEEKAAAKQVOTEKTTADTAIDGAHTNAEVEAAKNAEIAKIE 1133  
QY 416 APRDELLAKDK-----AFLQAE-----KDIADLGALYSQNFQLYKRVVEFR 457  
Db 1134 AIQPATTTKDNAKQAIATKANERKTAIAQTQDITAEETAAANAVNDNAVTOANNIEAAN 1193  
QY 458 SILRLDNADTTGNRNNTLVSGRSLIDPATNTL-----IVTDRSVIEKFRKLIDELDVP 512  
Db 1194 SONDYDQAKTIGES-----IDQVTPVKKKATAVTDKN----- 1228  
QY 513 AQQVMIEARIVEAADGFSRDLGVKFGA-TGKKCKKNDSAFSGWGVNSFGGDDKRWGASTK 571  
Db 1229 -----NITATD-----DNGVDTAKDAGKNSIQSTQAPATA-----VKSNAKND 1266  
QY 572 INLPITAAANSISLVRAISSGALNLSASESL-SKTKTLANPRVL-----TONRKE 622  
Db 1267 VQAVTTQNAID-----NITGATTEERNAKDLVLKAKKAYODILNAQTNDVTQIKQ 1322  
QY 623 AKIE-SGYEIPFTVTSIANGSSSTNTELKXAVLGLVTPNITPDGQIIMTVKINKDSPAQ 681  
Db 1323 AVADVQGITADTTIKDVAKDELATKAREQKALIAQTA-----DATTEEKEQANQOVDAE 1376  
QY 682 CASGNQTLICISTKMLNTQAMVENGGLIVGGIYEDNGNLTQKVPLLGDIPIVGNLFT 741  
Db 1377 LTQGNQNI-----ENASQI-----DDVNTAKDNAIQADIPQAS---T 1411  
QY 742 RGKKTDRRELLFITPRIMGTAGNS 766  
Db 1412 DVKTNAAREALLTEMQNKITEILNNN 1436

RESULT 22

US-10-092-880-9  
; Sequence 9, Application US/10092880  
; Publication No. US20020164354A1  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J.  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE  
; FILE REFERENCE: HAEMOPHILUS  
; CURRENT APPLICATION NUMBER: US/10/092,880  
; CURRENT FILING DATE: 2002-03-08  
; PRIOR APPLICATION NUMBER: 09/155,614  
; PRIOR FILING DATE: 1998-09-30  
; PRIOR APPLICATION NUMBER: 08/617,697  
; PRIOR FILING DATE: 1996-04-01  
; PRIOR APPLICATION NUMBER: PCT/US97/04707  
; PRIOR FILING DATE: 1997-04-01  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 1599





; Sequence 38, Application US/0928457  
; Patent No. US20020164603A1  
; GENERAL INFORMATION:  
; APPLICANT: DNA, specific proteins and peptides  
; TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method  
; TITLE OF INVENTION: for obtaining them and their biological application.  
; NUMBER OF SEQUENCES: 99  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30 (OEB)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/928,457  
; FILING DATE: 2001-08-14  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/214,759  
; FILING DATE: 199-12-10  
; INFORMATION FOR SEQ ID NO: 38:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1981 acid residues  
; TYPE: acid amin, s  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1..1981  
; US-09-928-457-38

Query Match 3.8%; Score 147.5; DB 10; Length 1981;  
Best Local Similarity 18.7%; Pred. No. 0.028;  
Matches 188; Conservative 119; Mismatches 342; Indels 355; Gaps 40;  
QY 1 MNTKLTIIISGLFVATAAFQTASAGNITDIKVSLLPNKQKIV--KVSFDEIYN---PTG 55  
DB 51 LKTLTKLVCSL-VLSGMVLPFAHAQITD---KSAKPNQVVLKNTGAPLVNIQTNG 106  
QY 56 FVTSSPARIALDFEQTGISMDQVLEADPLLSKISAAQNSRRARLVN-----104  
DB 107 RGLSHNRYTQFDVNDKGAVALND--RNNPFLVKS-----AQLINEVRGTASKLNG 157  
QY 105 -LNKPGQNTVEGRKNVWIFINESD-----TVSAP-----ARPAKAA 143  
DB 158 IVTVGGKADVIIANPNGITVNGGFKNVRGILITIGAPQTKDGAITGDFVROGLTVG 217  
QY 144 AAPAKQOQCR--TVQVRSIRIQLYPGKTTAAAPFTESVVSAPSPAKQQAASAKQ 201  
DB 218 AAGNDKGGADYTGVLARAVALQKLOKXNLAVSTGPKVDYASGEISAG-----267  
QY 202 QTAAPAKQOQTAAPAKQQAAPAKOTNIDFRKG---KNAGIIELAAALGFAGQPDISQHD 258  
DB 268 -TAAGTKPTIALDTAALGGMYADSTITLIANEKGVGVKNAGTLEA-----K 312  
QY 259 HIIVTLKNHTLPTTLQRSLSVDVDFKTPQKVTLRKLANDTOLIITTAGNMELVYKSAAPG 318  
DB 313 QLIVTSSG-----RIENSGRIATADGT-----EASPT 340  
QY 319 YFTFQVLPPKQNLKESGGVNNAPKTF-----TGRKISLD---FQDVEIR 358  
DB 341 YLSIETTEK-----GAAGTFISNGGRIESKGLLIVETGEDIISLRNAGVQNGSR 390  
QY 359 TILQILAKESGNIV-----ASDVNGKMTLSLKQVDPQADLWQ-----400  
DB 391 PATTVL--NAGHNLVIESKTNVNAKSANUSAGRTIIN-----DATIOAGSSV 438  
QY 401 -----ARNLDMRQQGNIVNAPRDELLAKDKAFLOAKEDIAIDLGLYLS 443  
DB 439 YSSTKGDTELGENTRIIAENVTVLSNIGSIAA---VIEAKDTAHIESGKPLSLSTVA 495  
QY 444 QNFOLKYKNVEFRSI-----L 460

DB 496 SNIRLNNNGIKGGKOLALLADDNITAKTNTLNLTNPNLYVHTGKDLNVLNVDKDLSAASHL 555  
QY 461 RLDN--ADTTGNRNTLVSGR-----GSLVIDPATNTLIVTDTRSVI 499  
DB 556 KSDNAAHITGTSKTLTASDKMVEAGLLNVNTNLTNSGNLHIOAAKGNIOLRNTKUNA 615  
QY 500 EXFRKLIDELDVPAQ--VMIEARIVEAADG-----FSRDLGVKFGAT 540  
DB 616 AK-----ALETALQGNIVSDGLHAVSGADGHSVLLANGNADFTGHNTLTAKADYNAGSV 669  
QY 541 GKXKLKND-----TSAFGNGVNSGFG-----GDDKWAETKINLPI 576  
DB 670 GGRLLKADNTNITSSSGDITLVAGNGIOLGDKQKNSINGKHISIKNGGNADLK--NLNV 728  
QY 577 TAAANSISLVRAISSGALNLSASELSKTKTLANPRVLTONRKEA-----KI 625  
DB 729 HAKSGALNI---HSDRALSIENKLESTHNAQHERVTLNQVDAYAHRLHSITGSOI 785  
QY 626 ESGYEIPFTVTSIANG-----GSSNTTELKCAVLGLTVTPNITPDGOI---IMTVK 673  
DB 786 WONDKLPSANKLVANGVIALNARYSQIADNTTLRAGAINLTAGTALVKGHNINWSTVSTK 845  
QY 674 INKDS-----AQCASGNQITLCISTKNLNTQAMVNGGTLIVGGIYEEDNGNTL 723  
DB 846 TLEDNAELKPLAGRLNIEAGSGTLTEPANRISAHITLSIKTGGKLLLSA-----KGGNAG 901  
QY 724 TKVPLLGDIPVIGNLFKTRGKKTDRRELLIFITPRIMGTAGNSL 767  
DB 902 APSAQVSSLEAKGNIRLVGTG-ETDURGSKI-----TAGKNL 936

RESULT 25  
US-10-066-551-1  
; Sequence 1, Application US/10066551  
; Publication No. US20030100071A1  
; GENERAL INFORMATION:  
; APPLICANT: Apicella, M. A.  
; APPLICANT: Edwards, J. L.  
; APPLICANT: Gibson, B. W.  
; APPLICANT: Scheffler, K.  
; APPLICANT: Brown, E.  
; TITLE OF INVENTION: Vaccine and compositions for the  
; TITLE OF INVENTION: prevention and treatment of Neisserial infections  
; FILE REFERENCE: 875.045051  
; CURRENT APPLICATION NUMBER: US/10/066,551  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: US 60/344,452  
; PRIOR FILING DATE: 2001-10-23  
; PRIOR APPLICATION NUMBER: US 60/310,356  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: US 60/266,070  
; PRIOR FILING DATE: 2001-01-31  
; NUMBER OF SEQ ID NOS: 12  
; SEQ ID NO 1  
; LENGTH: 2015  
; TYPE: PRT  
; ORGANISM: Neisseria gonorrhoeae  
; US-10-066-551-1

Query Match 3.8%; Score 147.5; DB 15; Length 2015;  
Best Local Similarity 18.7%; Pred. No. 0.028;  
Matches 188; Conservative 119; Mismatches 342; Indels 355; Gaps 40;  
QY 1 MNTKLTIIISGLFVATAAFQTASAGNITDIKVSLLPNKQKIV--KVSFDEIYN---PTG 55  
DB 51 LKTLTKLVCSL-VLSGMVLPFAHAQITD---KSAKPNQVVLKNTGAPLVNIQTNG 106  
QY 56 FVTSSPARIALDFEQTGISMDQVLEADPLLSKISAAQNSRRARLVN-----104  
DB 107 RGLSHNRYTQFDVNDKGAVALND--RNNPFLVKS-----AQLINEVRGTASKLNG 157  
QY 105 -LNKPGQNTVEGRKNVWIFINESD-----TVSAP-----ARPAKAA 143

Db 158. IVTGGQKADVIANPHTVNGGGFKNVGRGILTIGAPQIGKDGALTGFQVRRGTITVG 217  
Qy 144 AAPAKQOQCR--TVYQVRSIRIOTLYPGKTTAAAPFTESVVSAPPSPAKQQAASAKQ 201  
Db 218 AACWNDKGGADYTGVLARVALQKLGKKNLAVSTGPKDYASGEISAG-----267  
Qy 202 QTAAPAKQOQTAAPAKQQAAPAKQTNIDFRKDG---KNAGIIEALALGFAQGPDISQOHD 258  
Db 268 -TAAGTPTTIALDTAALGGMVADSIITLIANEKGVGVKNAGTLEAA-----K 312  
Qy 259 HIIVTLKXNHTLPTTLQRLSLVDAPKTPQVKVTLKRLANDTQLIITTAGNWLNVKSAAPG 318  
Db 313 QLIIVTSSG-----RIENSRIATTADGT-----EASPT 340  
Qy 319 YFTFQVLPKQNLSSGGVNNAPKTF-----TGKISLD-----FQDVEIR 358  
Db 341 YLSIETTEK-----GAAGTFISNGGRIBSKGLLVITETGEBDISLRNGAVVONGSR 390  
Qy 359 TILQILAKESGMNIV-----ASDVNGKMTLSLKDVPWQALDLVWQ-----400  
Db 391 PATTVL--NAGHNLVIESKTNVNNAKSANSLSAGRTIIN-----DATIOAGSSV 438  
Qy 401 -----ARNLDMRQOQNIVNIAPRDELLAKDKAFLOAEKDIADLQALYS 443  
Db 439 YSSTKGTDELGENTPRIIAENVTVLSNGSIGSAA--VIEAKDTAHIESGKPLSLETSTVA 495  
Qy 444 QNFOLKVKYVEEPSI-----L 460  
Db 496 SNILNNGNIGKGQALLADDNTAKTNLTPCNLYVHTGKDLNVLNVDKLSAASIHL 555  
Qy 461 RLDN-ADTGNRNLTLSGR-----GSVLIDPATNLTIVTDRSVI 499  
Db 556 KSDNAAHITGTSKLTASKDMGVREALVNTNLTNRNSGNLHIOAKGNIQLRNTKINA 615  
Qy 500 EKFRKLDELVDPAQ--VMIEARIVEAAG-----FSDI.GVKEGAT 540  
Db 616 AK-----ALETALQONI VSDGLUHVADGHVSLLANGNADFTGHNTLTAKADVNAGSV 669  
Qy 541 GKXKLKND-----TSAFGWVNSGFG-----GDDKWAETKINLPI 576  
Db 670 KGRLKADNTNITSSGDITLVAGNGIQLGDGKQNSINGKHISIKNNGNADIK.NLNV 728  
Qy 577 TAAANSISLVRAISSGALNELSSELSKTKTLANPRVLTQNRKEA-----KI 625  
Db 729 HAKSGALNI--HSDRALSIENTKLESHTNHLNAQHERVTLNQVDAYAHRLHSITGSQI 785  
Qy 626 ESGYEIPFTVTSIANG-----GSSTNTEKKAVLGLTVTPNITPDQOI---IMTVK 673  
Db 786 WQNDKLPKLVANGVLNARNARYSQIADNTTLRAGAINLTAGTALVKRGNINWSTVSK 845  
Qy 674 INKDS-----ACQASGNQITLICSTKNLNTQAMVENGTLIVGSIYEDNGNTL 723  
Db 846 TLEDNAELKPLAGRLNIEAGSGTLTIEPANRISAHDTLSIKTGKLLLSA---KGNAG 901  
Qy 724 TKVPLLDGIPVIGLNFKTRGKKTDRRELLIFITPRIMGTAGNSL 767  
Db 902 APSQVSSLEAKGNIRLVGT-EIDLROSKI-----TAGKNL 936

RESULT 26

US-09-815-242-5816  
; Sequence 5816, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haeelbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: PastSeq for Windows Version 4.0  
; SEQ ID NO 5816  
; LENGTH: 2478  
; TYPE: PRF  
; ORGANISM: Staphylococcus aureus  
; US-09-815-242-5816

Query Match 3.8%; Score 147.5; DB 9; Length 2478;  
Best Local Similarity 18.6%; Pred. No. 0.039;  
Matches 154; Conservative 114; Mismatches 300; Indels 261; Gaps 35;

Qy 6 TKIISGLFVATA-----AFQTASAGNI-----TDIKVSSLPNKQIKVSPDKIVNP- 53  
Db 725 TKAIKDIDAATTNAQVEAIKTKAINDINQTTTATTAKAAALEFDEVVQAQIDQAPLND 784  
Qy 54 -----TGFTVSSPARIALDFEQT-----GISMDQOVLEYADPLLSKISA 92  
Db 785 TTNEEVAEAIERINAAKVSGVKAIEATTTAQDLERVKNEEISKIENITDSTQTKMDAYNE 844  
Qy 93 AQNSRRARLVNLNLPKGVNTVGRGNKVMIFINESDDTVSAPARPAVKAAPAAKQOQC 152  
Db 845 VQKATARKAQAQATVSNATNEEVA-----EADAADVA-----AQKQL 882  
Qy 153 RTVYQVRSIRIOTLYPGKTTAAAPFTESVVSAPPSPAKQQAASAKQOQTAAPAKQ-- 210  
Db 883 HDIQVVKVS-----KQEVADTKSKVLKDKNATQTOAKVKPAADTEVENAYNTRKQEI 933  
Qy 211 ---TAAPAKQQAAPAKQTNIDFRK-----DGNAGIIEALALGFAQGPDISQOHDH 260  
Db 934 QNSNASTTEEKQAA---YELDTKKQEAETNLDAAVT-----NSDVTAKONS 978  
Qy 261 IVTLKXNHTLPTTLQRLSLVDAPF--KTPVQKVTLKRNLNDT-----QLIITTAG 306  
Db 979 IAAI-NOVQAATTKKSDAKABIAQKASERKTAIEAMNDSTTEEQQAAXDKVDQAVVT--A 1035  
Qy 307 NWEVLNKSAPGCFYFQVLFPKQNLSSGGVNNAPKTFTRKISLDFQDVEIR-TILOILA 365  
Db 1036 NADIDNAAA-----NNDVDNA-KTTNEATIAATPDANVVKPAKQAI 1077  
Qy 366 KESGMNIVASDSVNG-----KVTLSLKDVPWQALDLVWQARNLDMRQOQNIVN-- 414  
Db 1078 DKVQAGETAIDGNGSGTTEEKAAAQOQVTEKTTADAADAAHTNAEVEAAKKAIAKIE 1137  
Qy 415 -----IAPRELLAKOKAFLOAEKDIADLQALYSQNL 448  
Db 1138 AIQPATTTKDNKAKEAIATKANERKTAIAQTDITABEIAAANADV-----NAVTOAN-- 1190  
Qy 449 KYKNVEEFSILRLONADTTGNRNTLVSGRGSVLIDPATNLTIVTDRSVIEKFRKLIDE 508  
Db 1191 --SNIEANSQNDVDOAKTTGENS-----IDQVTPV-----1220  
Qy 509 LDVPAQQVMIARIVEAADGFSRDLGVKFGATGKKLKNXDTSAFGWGVNSGFGGDDKWA 568  
Db 1221 ----NKKATARNEITAILNKNLQEIQTATPDATBEKQAADEA-----NTENG-----1264

QY 569 ETKINLPITAAANS--ISLVRAISSGALNLELSASELSKTKYTLANPRVLTQNRKEAKIE 626  
Db 1265 --KANQAISAAATNAQVDEAKANAEEAII--AVTPKVV--KKQAARD 1305  
QY 627 SGYEIPFTVTSIANGSGSSTNTEKAV--LGLTVTP--NI---TPDQOIIMTVKINKD 677  
Db 1306 EIDQLQATQTNVINNDQNNATTEKEAAIQQLATAVTDKNNITAAATDDNGVDQADAGN 1365  
QY 678 S-----PAQCASGNQTLICSTKNLTQAMVENGTL--IVGGIYEEDN 719  
Db 1366 SIQSTQATATAVKS-----AKNDVDQAVTTQNTQNAIDNTTGTATEKN 1407

RESULT 27

US-09-815-242-12967  
; Sequence 12967, Application US/09815242  
; Patent No. US20020361569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Caffr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12967  
; LENGTH: 2478  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12967

Query Match 3.8%; Score 147.5; DB 9; Length 2478;  
Best Local Similarity 18.6%; Pred. No. 0.039; Indels 261; Gaps 35;  
Matches 154; Conservative 114; Mismatches 300;

QY 6 TKIISGLVATA-----AFQTASAGNI-----TDIKVSSLPNKQIKVKSFDKEIYNP- 53  
Db 725 TKAIKDIDAATTNAQVEAIKTKAINDINOTTPATTAKAAALEEFDEVVQAQIDQAPLND 784  
QY 54 -----TGFTVSPARIALDFECT-----GISMDQOVLEYADPILLSKISA 92  
Db 785 TTNEEVAEAIERINAAKVSQVKAIEATTTAQDLERVRNKEIEISKIENITDSTQTKXDAYNE 844  
QY 93 AQNSRAELVLNLPFGQYNTVEVRGNKVWIFINESDDTVSAPARPAVKAAPAAPAKQCGC 152  
Db 845 VKQAATARKAQAATVSNATNEEVA-----EADAADVA-----AQKQGL 882  
QY 153 RTVYVRSIRIQTLVPGKTTAAAPTESVUSVSAFSPAKQQAASAKAQQTAAPAKQO-- 210  
Db 883 HDIQVVK-----KQEVADTKSKVLDKINAIQTOAKVKPAADTEVENAYNTRKQEI 933

QY 211 --TAAPAKQAAAAKAKOTNIDFRK-----DGKNAGIIELAALGFAGOPDISQHHDI 260  
Db 934 QNSNASITTEKQAA--YTELDTKKQEARWLDAAANT-----NSDVTITAKDNS 978  
QY 261 IVTLKGNHTLPITLQKSLDVADE--KTPVQKVTLKRANDT-----QLIITAG 306  
Db 979 IAAI-NQVQAATTTKSDAKABIAQKASERKTAIEAMNDSTITEEQAAKDKVDQAVVT--A 1035  
QY 307 NWEVLNKSAAFGYFTFQVLPKKQNLSEGGVNNAPKTFTRKISLDFODVEIR--TILQILA 365  
Db 1036 NADIDNAAA-----NNDVDNA-KTTNEATIAAITPDANVKPAKQAIA 1077  
QY 366 KESGMNIVASDSVNG-----KMTLSLKDVDPDQALDLVMOQARNLDMFQOGNIVN-- 414  
Db 1078 DKVOAQETAIIDGNGSITTEEXAAAKQVQTEKTTADAAIDAHAHTNAEVEAAKKAIAKIE 1137  
QY 415 -----IAPRDELLAKKAFLOAKEDADLGLALYSQNFQ 448  
Db 1138 AIQPATTTKDNKAKEAIAIKANERKTAIAQTODITAEETAAANADV-----NAVTOAN-- 1190  
QY 449 KYKNVEFRSILRLDNADTTGNRTLVSGRGSVIDPATNTLIVTDRSRVIEKFRKLIDE 508  
Db 1191 --SNIEAANSQNDVDQAKTTGENS-----IDQVTFTV----- 1220  
QY 509 LDVPAQQVMIETARIVEAADGFSRDLGVKFGATGKKLKNDDTSAFGMGVNSGFGGDDRWGA 568  
Db 1221 -----NKKATARNEITAILNNKLQEIQAOTPDATDEKQADAEA-----NTENG-- 1264  
QY 569 ETKINLPITAAANS--ISLVRAISSGALNLELSASELSKTKYTLANPRVLTQNRKEAKIE 626  
Db 1265 --KANQAISAAATNAQVDEAKANAEEAII--AVTPKVV--KKQAARD 1305  
QY 627 SGYEIPFTVTSIANGSGSSTNTEKAV--LGLTVTP--NI---TPDQOIIMTVKINKD 677  
Db 1306 EIDQLQATQTNVINNDQNNATTEKEAAIQQLATAVTDKNNITAAATDDNGVDQADAGN 1365  
QY 678 S-----PAQCASGNQTLICSTKNLTQAMVENGTL--IVGGIYEEDN 719  
Db 1366 SIQSTQATATAVKS-----AKNDVDQAVTTQNTQNAIDNTTGTATEKN 1407

RESULT 28

US-10-193-764-61  
; Sequence 61, Application US/10193764  
; Publication No. US2003013943A1  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH  
; FILE REFERENCE: 1038-1239MIS  
; CURRENT APPLICATION NUMBER: US/10/193,764  
; CURRENT FILING DATE: 2002-07-12  
; PRIOR APPLICATION NUMBER: 09/167,568  
; PRIOR FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 91  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 61  
; LENGTH: 1180  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-193-764-61

Query Match 3.8%; Score 146.5; DB 12; Length 1180;  
Best Local Similarity 20.8%; Pred. No. 0.014;  
Matches 101; Conservative 80; Mismatches 182; Indels 122; Gaps 27;

QY 294 LNNDTOLIIITAGNWEVLNKSAAFGYFTFQVLPKKQNLSEGGVNNAPKTFTRKISLDFQ 353  
Db 357 INNDLTL-NATGNISLLQVEGIDMIGGVAKK-NI-----TFTGNITFGSK 404



Query Match 3.7%; Score 143.5; DB 12; Length 1220;  
Best Local Similarity 21.1%; Pred. No. 0.027;  
Matches 103; Conservative 79; Mismatches 178; Indels 127; Gaps 27;

Qy 183 S-----VSAPFSPAKQAAASAKQOATAAPAKQOAAAPAKQTNIDF 230  
Db 853 AEDGVDAKYSYSLDAGSKLQREAGSDHQLTADDLAEVTTITITPAIATADSDNVSF 912  
Qy 231 RKDGKN-----AGITELALGAGQPDISQHDHIV-----262  
Db 913 EYNGKTKASEAEGIQATVNLGSGGVVALTSADVVDVNDGVDAKYSYQSLDAGKAKLOA 972  
Qy 263 -TLKQHTL-----PTTLQRSLDVA---DFKTPVO-----KYTLKRLNND 297  
Db 973 ATGNVYQLTADDLKVTGTITITPAITTVDSNDVSFEYDGKTKAGEAKGIQVTVK-LGET 1031  
Qy 298 TQLIITAGNWLNVKSAAPGYFTFOVLPK-KONLESGGVNAPKFTTGRKISLDFQVE 356  
Db 1032 EKTVDLTSADIVVANDVNDVAGQYSYQSLDAGKAKLOA-----TGNVQLTADD-- 1080  
Qy 357 IRTILQILAKESGMNIVASDSVNGKMTLSLKDVPMD-----QALDLVMQ---AR 402  
Db 1081 -----LAKVTG-TITITPAVT---TADSDNVSEYDGKTKASEAKGIQVIVKLGETEK 1129  
Qy 403 NLDMRQGNINVIAPDELLAKDKAFLOAEKDIA-DLGALY-----SONFOL-----KYKXVEEFRS 458  
Db 1130 TVDLTSADIVV---ANDVDNAGYSYQSLDAGKAKLOAATGNNYQL-----1172  
Qy 463 DNADTTGNRTLVSGRGSVLIDPATNTLIVTDTFRSVIERFKRLIDELDPVPAQVMIIE--- 519  
Db 1173 -----TADDLAKITGTITITPAVT---ADSNVSEYDGKTKASEARGIQATVKLGE 1222  
Qy 520 -----ARIVEAADGFSR-DLVKFGATGKKLKNUTSAPFGVNGSGFGGDD---KWG 567  
Db 1223 NGKTVALTADIIVVNDGVNDVAGQYDKLSAAGTKLQAT-----GTNYQFKKEDTLKLG 1277  
Qy 568 AETKINLPITAAA--NSISL-----VRAISSGAL--NLELS-----ASELSKTKTLANPRV 615  
Db 1278 GTVTIT-PATALADLNDVSFSDGQYKASQAHDLTANIKLGTAKVSVHLNATDILVTDDG 1336  
Qy 616 LTQNRKEAKIESGYEIPFTVTSIANG-----GSSTNTELKKAUL-GLTVPNTITPDGQI 668  
Db 1337 VGVQYQYKLD-----ANGIAKLQASGNDYQDAKVLAGLTGTITIKP---V 1381  
Qy 669 IMTVKINKDS-----PACASGNITILCI-----STKNLNTQ-AMVENGQTL-----709  
Db 1382 TGAVTVNDTSFVYDGHGKASAAAGLOASLVLPOAEAKATQLTREILVINDGTAAGTYR 1441  
Qy 710 -----IVGGIYEDNG-----NTLTKVPLLGDIPIVGNLFTRG 743  
Db 1442 YRLSOTGIAKLQKAVGKNYELDQDELAGLTGTITITPLTVNATVNHGQFOYNG 1494

RESULT 31  
US-10-193-764-28  
; Sequence 28, Application US/10193764  
; Publication No. US20030133943A1  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH  
; FILE REFERENCE: 1038-1239MIS  
; CURRENT APPLICATION NUMBER: US/10/193,764  
; CURRENT FILING DATE: 2002-07-12  
; PRIOR FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 91  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 28  
; LENGTH: 1220  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-193-764-28

Query Match 3.7%; Score 143.5; DB 12; Length 1220;  
Best Local Similarity 21.1%; Pred. No. 0.027;  
Matches 103; Conservative 79; Mismatches 178; Indels 127; Gaps 27;

Qy 294 LNNDTQLIITAGNWLNVKSAAPGYFTFOVLPKKONLESGGVNAPKFTTGRKISLDFQ 353  
Db 353 INNDLTL-NATGGNISLLQVEGIDGMIGKGVAKK-NI-----TFAGGNITFGSK 400  
Qy 354 DVEIRTLQI-----LAKESGMNIVASDSVNGKMTLSLKDVPMDQALDLVMOARNLDMRQ 408  
Db 401 ----KATELEGNATINNANVTILIGSDFNHOKPLTIKK-----DVIINSGL--TA 447  
Qy 409 QGNIVNIAPDELLAKDKAFLOAEKDIA-DLGALY-----SONFOL-----KYKXVEEFRS 458  
Db 448 GGNVINI--NGNLTAVNGANLKAITNFTNVGGLFDKNKGSNISIAARGGAKFKDINNNTSS 505  
Qy 459 ILRLDNADTTGNRTLVSGRGSVLIDPATNTLIVTDTFRSVIERFKRLIDELDPVPAQVMI 518  
Db 506 LNIITNSDIT--YRTIIEGN-----ITNKAGDNLINDKGNA-----EIQIGNISQK 551  
Qy 519 EARIVEAADGFSRDLGVKFGATGKKLKNUTSAPFGVNGSGFGGDDKXGAETKINLPITA 578  
Db 552 EGNLTISD-----KINITNOITIKK-----GVNK-----EDSDSTANNANLTI 591  
Qy 579 AANSISLVRAIS-SCALNLELSASESI-----SKTKTLANPRVLTONR-KEAKIESG-Y 629  
Db 592 KTKELQLTGDLNISGFKAEITAKEGADLIIGNSDNNNANAKKVTFNQVXDSKISAGSH 651  
Qy 630 BIPP---TVTSIANGSGSTN---TELKKAVLGLTVPNTITPDGQIIMTVKINKDSPAQ 682  
Db 652 NVTLSKVTESNGNDAESNNGDSTLTINAKNVTVANNIT-----692  
Qy 683 ASGNQTLICISTKNLNTQAMVENGQTL-----IVGGIYEDNGNTLTKVPLLG 731  
Db 693 --SHKTVNITASENVTTKAGTTINATIGSVETAKTGDKIGGI-ESNSGN--VNITASGD 747  
Qy 732 IPVIGNL 738  
Db 748 TLNVSNI 754

RESULT 32  
US-10-193-764-26  
; Sequence 26, Application US/10193764  
; Publication No. US20030133943A1  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH  
; FILE REFERENCE: 1038-1239MIS  
; CURRENT APPLICATION NUMBER: US/10/193,764  
; CURRENT FILING DATE: 2002-07-12  
; PRIOR FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 91  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 26  
; LENGTH: 1226  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-193-764-26

Query Match 3.7%; Score 143.5; DB 12; Length 1226;  
Best Local Similarity 21.1%; Pred. No. 0.027;  
Matches 103; Conservative 79; Mismatches 178; Indels 127; Gaps 27;

Qy 294 LNNDTQLIITAGNWLNVKSAAPGYFTFOVLPKKONLESGGVNAPKFTTGRKISLDFQ 353  
Db 359 INNDLTL-NATGGNISLLQVEGIDGMIGKGVAKK-NI-----TFAGGNITFGSK 406  
Qy 354 DVEIRTLQI-----LAKESGMNIVASDSVNGKMTLSLKDVPMDQALDLVMOARNLDMRQ 408



```
Query Match 3.7%; Score 143; DB 9; Length 2076;
Best Local Similarity 18.4%; Pred. No. 0.069;
Matches 132; Conservative 97; Mismatches 290; Indels 200; Gaps 28;

QY 51 VNPTGVTSSPARIALDFEQTGISMDOQ-----VLEYADPLLSKISAQNSR 98
Db 879 INPI-----APVTVVKQAAARDASHDAQOAHIAEINANPDATQEEEOAIDKVNAAVTAAN 933
QY 99 ARVLNLNKPGQYNTVRGNKVMIFINESDDTVSAPAPVAKA-APAAPAKQGGCTVYQ 157
Db 934 TN-ILNAN-----TNADVEQVKTNAIOGQIAITPATKVTDAKNAIDK 975
QY 158 VRSIRIOTLPYKTTAAAPFTESVVSAPFSPAKQQAASAKQOATAAPAKQOATAAPAKQ 217
Db 976 SAETQHTNFIENN-----NDATLEEQAAQQLLDQAVATAKQINNAADTNOEVAQAKD 1027
QY 218 QAA-----APAKQTNIDFR---KDGKNAGIIEAALGFAGQPDISQQHDIHIVTLKNT 268
Db 1028 QGTQNIWVQPATQVKTDRNVVNDKAREAITNATTGATREE-KOEAINRVNTLKNRA 1086
QY 269 LPTTLQRSLOVADFPTQVQVTKLRNNDTQLIITTAGNWLKNSAAPGYFTFVLPKK 328
Db 1087 L-----TDIGVSTTAMVNSIR---DDAVNOIGAVQPHVTKQTATGVNLNDLATAKK 1135
QY 329 QNLESGGVNNAKPTFTGRKISLDFQDVEIRTILOIL-----AKESGM----- 370
Db 1136 QEIN-----QNTNATTEEKQVALNQVDQELATANNINQADTNAEVDQAQOLGTKAINAIQ 1191
QY 371 NIV-----RNLDMRQCGNIVNIAPRDELLAKDKAFLOAEKDI-----ADLGALYSQNF 446
Db 1192 PNIVKKAALAQINOHYNKALAEINATPDATNDEKNAINTLNQDRQQAIESIKQANTNA 1251
QY 447 QLKYNKVEFRSILRLDNADTTGNR--NTLVSGRGSVLIDPATNTLIVTTRSVIEKFRK 504
Db 1312 QLKQQAQINQINQNTNDQVDTTITNQAVNAIDNVEAEVVKPKA-----IADIEKAVKEKQ 1367
QY 505 LIDELDPQAQVMIIEARIVEAADGFSRDLGVKFGATGKKLKNDTSAFGWVNSGFGGDD 564
Db 1368 QID-----NSLSDTNEKEVASQALAKEKEK-----ALAAID 1399
QY 565 KWGAETKINLPITAAANSISLVRAISSGALNLELSASELSKTKTLANPRVLTO-NRKEA 623
Db 1400 QAQTNSQVN---QAATNGVSAIKII-----QPETKVKPAAREKINQKANELRA 1444
QY 624 KIESGYE-----IPFTVTSIANGSGSSTNTELKAV-LGLTVPNTPTD 665
Db 1445 KINQKQKATAEERQVALDKINEFVNQAMTDITN--NRTNQVQDDTTSQALDSIALVTPD 1501

RESULT 35
US-09-815-242-12913
Sequence 12913, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haseibeck, Robert
APPLICANT: Chisen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
```

```
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12913
LENGTH: 2186
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12913

Query Match 3.7%; Score 143; DB 9; Length 2186;
Best Local Similarity 18.4%; Pred. No. 0.075;
Matches 132; Conservative 97; Mismatches 290; Indels 200; Gaps 28;

QY 51 VNPTGVTSSPARIALDFEQTGISMDOQ-----VLEYADPLLSKISAQNSR 98
Db 879 INPI-----APVTVVKQAAARDASHDAQOAHIAEINANPDATQEEEOAIDKVNAAVTAAN 933
QY 99 ARVLNLNKPGQYNTVRGNKVMIFINESDDTVSAPAPVAKA-APAAPAKQGGCTVYQ 157
Db 934 TN-ILNAN-----TNADVEQVKTNAIOGQIAITPATKVTDAKNAIDK 975
QY 158 VRSIRIOTLPYKTTAAAPFTESVVSAPFSPAKQQAASAKQOATAAPAKQOATAAPAKQ 217
Db 976 SAETQHTNFIENN-----NDATLEEQAAQQLLDQAVATAKQINNAADTNOEVAQAKD 1027
QY 218 QAA-----APAKQTNIDFR---KDGKNAGIIEAALGFAGQPDISQQHDIHIVTLKNT 268
Db 1028 QGTQNIWVQPATQVKTDRNVVNDKAREAITNATTGATREE-KOEAINRVNTLKNRA 1086
QY 269 LPTTLQRSLOVADFPTQVQVTKLRNNDTQLIITTAGNWLKNSAAPGYFTFVLPKK 328
Db 1087 L-----TDIGVSTTAMVNSIR---DDAVNOIGAVQPHVTKQTATGVNLNDLATAKK 1135
QY 329 QNLESGGVNNAKPTFTGRKISLDFQDVEIRTILOIL-----AKESGM----- 370
Db 1136 QEIN-----QNTNATTEEKQVALNQVDQELATANNINQADTNAEVDQAQOLGTKAINAIQ 1191
QY 371 NIV-----RNLDMRQCGNIVNIAPRDELLAKDKAFLOAEKDI-----ADLGALYSQNF 446
Db 1252 EVDQAATVAENNIDAVQDVVVKQAARDKITAEVAKRIEAVKQTPNATDEKQAAVNOIN 1311
QY 447 QLKYNKVEFRSILRLDNADTTGNR--NTLVSGRGSVLIDPATNTLIVTTRSVIEKFRK 504
Db 1312 QLKQQAQINQINQNTNDQVDTTITNQAVNAIDNVEAEVVKPKA-----IADIEKAVKEKQ 1367
QY 505 LIDELDPQAQVMIIEARIVEAADGFSRDLGVKFGATGKKLKNDTSAFGWVNSGFGGDD 564
Db 1368 QID-----NSLSDTNEKEVASQALAKEKEK-----ALAAID 1399
QY 565 KWGAETKINLPITAAANSISLVRAISSGALNLELSASELSKTKTLANPRVLTO-NRKEA 623
Db 1400 QAQTNSQVN---QAATNGVSAIKII-----QPETKVKPAAREKINQKANELRA 1444
QY 624 KIESGYE-----IPFTVTSIANGSGSSTNTELKAV-LGLTVPNTPTD 665
Db 1445 KINQKQKATAEERQVALDKINEFVNQAMTDITN--NRTNQVQDDTTSQALDSIALVTPD 1501
```

## RESULT 36

US-10-193-764-35  
; Sequence 35, Application US/10193764  
; Publication No. US20030133943A1  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH  
; FILE REFERENCE: 1038-1239MIS  
; CURRENT APPLICATION NUMBER: US/10/193,764  
; PRIOR FILING DATE: 2002-07-12  
; PRIOR APPLICATION NUMBER: 09/167,568  
; PRIOR FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 91  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 35  
; LENGTH: 915  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-193-764-35

Query Match 3.7%; Score 142.5; DB 12; Length 915;  
Best Local Similarity 21.3%; Pred. No. 0.02;  
Matches 138; Conservative 80; Mismatches 209; Indels 221; Gaps 34;

|    |     |               |              |              |               |             |            |     |
|----|-----|---------------|--------------|--------------|---------------|-------------|------------|-----|
| QY | 227 | NIDFRKQGN     | ---AGII---   | ELAALGFAGQPD | ISQQDHIIIVTLK | NHT         | -----      | 268 |
| DB | 5   | NHFNASSNYQTYG | VIIBSQNFSAGG | SSLKFKXSEGS  | THAAFTIKNDL   | ILNATGNISL  | 64         |     |
| QY | 269 | -----LPTTLQ   | RSLS         | -----        | DVADFKTPVQ    | -----       | 297        |     |
| DB | 65  | NQVAGIDSNLKS  | LANKNITFE    | GGNITLAADKPK | IEIKGNITVKE   | GANVILRSANY | 124        |     |
| QY | 298 | TQLI          | -----        | ITTAGNWLKNS  | -----         | AAPGYFTFQVL | PKQNLES    | 345 |
| DB | 125 | KGALSIRGNVT   | KNLTVGSA     | INTEKNTVEG   | SAKFLANPNYS   | FNVSGLFNO   | 181        |     |
| QY | 346 | RKISLD        | -----        | FQDVEIRI     | LOILAKESGM    | NIVASDS     | -----      | 395 |
| DB | 182 | NISIAKGAIF    | KPDIENTGSL   | NITTK        | -----         | SDSNHHTI    | IKGNITNR   | 226 |
| QY | 396 | DLVMQARNLDM   | ROQNI        | -----        | VNIAPRDELLA   | KDKAFLOAEK  | DIADLGALY  | 442 |
| DB | 227 | NITNNGDTEI    | QIGGNISOK    | EGNLTIS      | SDKNITERIT    | IK          | -----      | 280 |
| QY | 443 | SONFOLKYK     | -----        | NVEEPR       | --SILRLDNAD   | TGNRNTLV    | SGRGSVLID  | 492 |
| DB | 281 | SANLTIKTEL    | KLTNDLNI     | SGFNKAEI     | TAKDNSNLT     | IGDN        | -----      | 325 |
| QY | 493 | TDRSVIEKRL    | ID-ELDVP     | QAQVMIEAR    | IVEAADGFS     | RDLG        | -----      | 536 |
| DB | 326 | TDARKV        | --TFSNVK     | DSKISASDH    | NVILNSKVT     | SGTSDT      | DEGGNNNTGL | 383 |
| QY | 537 | FGATGKKKL     | -----        | KNDTSAFG     | WVNSGFG       | -----       | GDDKWAETKI | 582 |
| DB | 384 | NNITSHKVT     | NITASENV     | TTKAGTTIN    | ATTGSEVTA     | KTDGDKG     | ESNSGNVNI  | 443 |
| QY | 583 | ISLVR         | -----        | AISGALN      | LESASELS      | SKTKTLAN    | PRVLTONRK  | 631 |
| DB | 444 | LNVSNITGQ     | NVTAAASGA    | -----        | VTTTKGST      | INATTGNANI  | TTKTEINGE  | 498 |
| QY | 632 | PFT           | -----        | VTSIANG      | SSSTNELK      | KA          | VLGLTVP    | 676 |
| DB | 499 | NITASGNL      | NVSNITG      | NVTANS       | GAITTE        | -----       | GSTI--NATT | 550 |
| QY | 677 | DSPAQACAS     | GNQTLICIS    | TKNLNTQ      | AMVENG        | GTLVGGI     | YEEDNGNTLT | 724 |
| DB | 551 | NGKVES        | SSGSVT       | -----        | LIATGQTL      | AVGNI       | -----      | 581 |

## RESULT 37

US-10-193-764-37  
; Sequence 37, Application US/10193764  
; Publication No. US20030133943A1  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH  
; FILE REFERENCE: 1038-1239MIS  
; CURRENT APPLICATION NUMBER: US/10/193,764  
; CURRENT FILING DATE: 2002-07-12  
; PRIOR FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 91  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 37  
; LENGTH: 1222  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-193-764-37

Query Match 3.7%; Score 141.5; DB 12; Length 1222;  
Best Local Similarity 21.4%; Pred. No. 0.03;  
Matches 135; Conservative 79; Mismatches 201; Indels 217; Gaps 33;

|    |     |             |               |               |              |             |                 |     |
|----|-----|-------------|---------------|---------------|--------------|-------------|-----------------|-----|
| QY | 239 | IIELAALGFAG | QPD           | -----         | SQCHDHIIIVTL | KNHT        | -----           | 277 |
| DB | 328 | IIESQNF     | SASGSSLKFK    | XSEGS         | THAAFTIKNDL  | ILNATGNISL  | NQVAGIDSNL      | 387 |
| QY | 278 | -----       | DVADFKTPVQ    | -----         | KVTLKRLN     | --NDTQLI    | -----           | 308 |
| DB | 388 | NKNTFE      | GGNITLAADKPK  | IEIKGNITVKE   | GANVILRSANY  | GNDKLSALSIR | GNVTNKG         | 447 |
| QY | 309 | ELVNS       | -----         | AAPGYFTFQVL   | PKQNLES      | GMNNA       | PFTTGRKISLD     | 356 |
| DB | 448 | TVTGSA      | INIEKNLTV     | EGSAKFLANPNYS | FNVSGLFNO    | GKS         | -----           | 503 |
| QY | 357 | IRTLQILAK   | ESGMNIVASDS   | -----         | VNGKMTLSL    | KDVPD       | WDQALDVMQARNLDM | 411 |
| DB | 504 | NTGSLNITTK  | -----         | SDSNHHTI      | IKGNITNR     | KGD         | -----           | 549 |
| QY | 412 | I           | -----         | VNIAPRDELLA   | KDKAFLOAEK   | DIADLGALY   | SONFOLKYK       | 451 |
| DB | 550 | ISQEGNLT    | ISSDKVNITERIT | IK            | -----        | GVNGD       | NSDSNEATSANLT   | 603 |
| QY | 452 | NVEEPR      | --SILRLDNAD   | TGNRNTLV      | SGRGSVLID    | PATNTL      | IVTDRSVIEK      | 507 |
| DB | 604 | LNISGF      | NKAEITAKD     | NSNLTIGDN     | -----        | SDAGN       | -----           | 646 |
| QY | 508 | ELDVP       | QAQVMIEAR     | IVEAADGFS     | RDLG         | -----       | VKFGATGKKKL     | 547 |
| DB | 647 | KISASDH     | NVILNSKVT     | SGTSDT        | DEGGNNNTG    | LTITAKNV    | VNNNITSHKVT     | 706 |
| QY | 548 | DTSAFG      | WVNSGFG       | -----         | GDDKWAETKI   | NLPITAAANS  | SILVR           | 590 |
| DB | 707 | VTTKAG      | TTINATTG      | SEVTA         | KTDGDKG      | ESNSGNVNI   | TASGDTLNV       | 766 |
| QY | 591 | SGALN       | LESASELS      | SKTKTLAN      | PRVLTONRK    | -----       | EAKIESGYEIPFT   | 634 |
| DB | 767 | SGA         | -----         | VTTTKGST      | INATTGNANI   | TTKTEINGE   | VKSASG          | 821 |
| QY | 635 | --VTSIANG   | SSSTNELK      | KA            | VLGLTVP      | NPITPD      | QCIIMTVK        | 692 |
| DB | 822 | QNVTVT      | ANS           | GAITTE        | -----        | GSTI--NATT  | GDNITTTQ        | 869 |
| QY | 693 | STKNL       | NTQAMV        | ENG           | GTLVGGI      | YEEDNGNTLT  | 724             |     |
| DB | 870 | -----       | LIATGQTL      | AVGNI         | -----        | SGDTVT      | 888             |     |

## RESULT 38

US-10-193-764-34



Sequence 34, Application US/10193764  
Publication No. US20030133943A1  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M.  
INVENTOR: Yang, Yan-Ping  
APPLICANT: Klein, Michel H.  
TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH MOLECULAR WEIGHT PROTEINS  
FILE REFERENCE: 1038-1239MIS  
CURRENT APPLICATION NUMBER: US/10/193,764  
PRIORITY FILING DATE: 2002-07-12  
PRIORITY FILING DATE: 09/167,568  
PRIORITY FILING DATE: 1998-10-07  
NUMBER OF SEQ ID NOS: 91  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 34  
LENGTH: 1228  
TYPE: PRT  
ORGANISM: Haemophilus influenzae  
US-10-193-764-34

Query Match 3.7%; Score 141.5; DB 12; Length 1228;  
Best Local Similarity 21.4%; Pred. No. 0.039;  
Matches 135; Conservative 79; Mismatches 201; Indels 217; Gaps 33;

QY 239 IIELAALGFAGOPDI---SQQDHIIIVTLKQHT-----LPTTLQRLS-- 277  
DB 334 IIESQNFASGSSSLKFKSEGSTHAFTIKNDLILNATGNSISLQVAGIDSNLKSLIA 393  
QY 278 -----DVADKPTPVQ-----KVLKRLN--NDQLI-----ITTAGNW 308  
DB 394 NKNITPEGGNITLADKKPIETKGNITVKEGANVILRSANYGNDKSALSIRGNVTKGNL 453  
QY 309 ELVNKS-----AAGYFTFQVLPKKQNLSEGGVNNAPKTTGKISLD-----FDVVE 356  
DB 454 TVTGSAINIEKNTVGSAAKFLANPNYSFNVSLFPNQKS-----NISAKGAIFKDI 509  
QY 357 IRTILOILAKESGMNIVASDS-----VNGKMTLSLKDVPWDQALDLVQARLNDNRQON 411  
DB 510 NTGSLNITTK-----SDSNHHTIKGNITNRKGD-----LNTTNGNDNTEIQIGN 555  
QY 412 I-----VNIAPDELLAKDKAFLQAEKDIALGALYSONFOLKVK----- 451  
DB 556 ISOKEGNLTSSDKVNITERITKA-----GVNGDSDSNEATSANLIKTKELKLTND 609  
QY 452 -NVEEER--SILRLDNADITGNRLTVSGRGVLDIPANTLIVTDRSVIEKFKLID- 507  
DB 610 LNISGFNKABITAKDNSNLTIQGN---SDAGN-----TDKKV--TFSNVKDS 652  
QY 508 ELVPAQOQVMEARIVEAAGFSDRLG-----VKFGATGKKKL-----KN 547  
DB 653 KISASDHNVTLNSKVTSGDSTEDCGNNITGLTITAKVTVNNITSHKTVNITAS 712  
QY 548 DTSAGFGVNSGFG-----GDDKMGAEKI-NLPITAAANSISLVR-----AIS 590  
DB 713 VTKAGTINATGSGVEVTAKTGDIKGBIESNSGNVNITASGDTLNVSNITQNVTVAAA 772  
QY 591 SGALNELSASELSKTKTLANPRVLQNRK--EAKIESGYEIPPT----- 634  
DB 773 SGA-----VTTKGTINATGNANITTKGEINCEVKSAG-NVNITASNTLNVSNITG 827  
QY 635 --VTSIANGSGSSTTELKXAVLGTVTPNITPPQOIIMTVKINDSPQAQASGNQITLCI 692  
DB 828 QNVTVTANSIAITTE-----GSTI--NATTGDANITTTGTNGINGKVESSSGVT---- 875  
QY 693 STKNLNTQAMVENGTLIVGGIYEEDNNTLT 724  
DB 876 -----LIATQOTLAVGNI-----SGDTVT 894

RESULT 39

US-09-797-862-33

; Sequence 33, Application US/09797862

Patent No. US20020102276A1  
GENERAL INFORMATION:  
APPLICANT: PEAK, IAN RICHARD ANSELM  
INVENTOR: JENNINGS, MICHAEL PAUL  
APPLICANT: MOXON, E. RICHARD  
TITLE OF INVENTION: NOVEL SURFACE ANTIGEN  
FILE REFERENCE: 065064/0134  
CURRENT APPLICATION NUMBER: US/09/797,862  
PRIORITY FILING DATE: 2001-05-03  
PRIORITY FILING DATE: 1998-12-14  
PRIORITY FILING DATE: 9726398.2  
PRIORITY FILING DATE: 1997-12-12  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 33  
LENGTH: 2353  
TYPE: PRT  
ORGANISM: Haemophilus influenzae  
US-09-797-862-33

Query Match 3.7%; Score 141; DB 10; Length 2353;  
Best Local Similarity 20.4%; Pred. No. 0.12;  
Matches 164; Conservative 105; Mismatches 338; Indels 198; Gaps 36;

QY 25 GNITDIKVSLLPNKKQKIVKVPKKEIVNPTGFTVSSPARIALD-----FEQTGISMDQQ 78  
DB 664 GDTIKLKVDN-QNTDNLVLTGNNGTAVTKGGFETVKTGATDARGKTVVKDANDADKK 722  
QY 79 VLEVADPLLSKISAAQNSRRLVNLNPKQCVNTEVRGNKVMIFINESDDTIVSAPARPA 138  
DB 723 VATVKDVATAINSAAFTVKTENLTTSIDEDNPTD-----NGKDALKAGDTLT 770  
QY 139 VKAAPAPAKQOQGRVYQV-RSRIOTLYPGKT-----TAAAPFTSVSVSVA 186  
DB 771 FKAGNLKVKRDKGNITFDLAKNLEVTAKVSDTLTIGGNTPTGTTATP-KVNITSTAD 829  
QY 187 PFSKQQAASAKAQOQTAAPAKQOQTAAPAKQQAAPAKQTNIDPRKDGKNAIIEALAG 246  
DB 830 GLNFAKETADAGSGNVYKGIATLTLEPSNAGKSSHVLDNVDAATKKSNAASIEDVLRAG 889  
QY 247 FAGQFDSIQQHDHI-----IVTLKNHTLPTLQRLSDVADFKTPVQKVTLKLNDTOLII 302  
DB 890 WNIQGN-GNNVDYVATVDTVNFDTDDSTGTTTIVTVTKADGKGADVKI-GAK-----TSVI- 942  
QY 303 TTAGNWLNVKSAAPGYFTFQVLPKKQNLSEGGVNNAPKTTGKISLDFOQVEI-RTIL 361  
DB 943 -----KDNH-----GKLFYTK-----DLKDANNGATVS 965  
QY 362 QILAKESGMNIVASDSVNGKMTLSLKDVPWDQAL-----DLVQARNLDMRQOQNVNIAP 417  
DB 966 EDDGKDTGTLVGTAKTVIDAVNKSQWRTGEGATAETGATAVNAGNAETVTSVNFKN 1025  
QY 418 RDELLAKDKAFLQAEKDIALGALYSONFOLKYNVEEFSILRLDNADITGNRLTVSG 477  
DB 1026 GNATTA-----TVSKONGNINVKYDVNVGGGLKIGDDKKIV-----ADTT-----TLTVT 1070  
QY 478 RGSVLIDPATNTLIVTDRSVIEKFKLI--DELDVPAQOQVMEARIVEAAGFSDR---D 532  
DB 1071 GKQVSVFAGANS-----VNNKKLVNAEGLATALLNLSWTAKADYADGESEGETD 1121  
QY 533 LGVKEFG-----ATGKK-KLKNDTSAFGV-----NSGFGGDKMGAEKIN 573  
DB 1122 QEVKAGDKVTFKAGKXVLKQSEKDFTVYSLODTLTGLTSITLGGTANGRNDTGVINKDG 1181  
QY 574 LPITAA-----ANSISLVR-AISSGALNIELSASELSKT-KTLANPRVLQNR 620  
DB 1182 LTIITLANGAAGTDSANGNTISVTKDGI-SAG--NKEITNVKSALKTYKDTQNTADETQDK 1239  
QY 621 K-EAKIESGYEIPPTVTSIANGSGSSTTELKXAVLGTVTPNIT-----PDQOI 668  
DB 1240 EFHAAVKNANEVEFVGKNGATVSAKTNNNGKH-----TVTIDVAEKVGDGLEKDTDGKI 1294

Search completed: December 9, 2003. 10:32:59

Search completed: December 9, 2003, 10:32:59  
Job time : 46 secs

Search completed: December 9, 2003, 10:32:59  
Job time : 46 secs

Search completed: December 9, 2003, 10:32:59  
Job time : 46 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 10:23:57 ; Search time 21 Seconds  
(without alignments)  
3521.606 Million cell updates/sec

Title: US-09-701-271A-2  
Perfect score: 3848  
Sequence: 1 MNTKLTKIISGLFVATAAFQ.....ELLFITPRINGTAGNSIRY 769

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9618682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID       | Description        |
|------------|--------|-------------|--------|----------|--------------------|
| 1          | 3666   | 95.3        | 761    | 2 A81985 | pilus secretin NVA |
| 2          | 2988   | 77.7        | 720    | 2 S70838 | pilQ protein - Nei |
| 3          | 2765.5 | 71.9        | 711    | 2 A37051 | outer membrane pro |
| 4          | 984.5  | 25.6        | 714    | 2 A83016 | type 4 fimbrial bi |
| 5          | 959.5  | 24.9        | 714    | 2 S37345 | pilQ protein - Pse |
| 6          | 907    | 23.6        | 637    | 2 D82816 | fimbrial assembly  |
| 7          | 809.5  | 21.0        | 578    | 2 D82053 | fimbrial assembly  |
| 8          | 574.5  | 14.9        | 412    | 2 A81000 | type II secretion  |
| 9          | 567.5  | 14.7        | 412    | 2 A91158 | probable transport |
| 10         | 567.5  | 14.7        | 412    | 2 G86003 | probable transport |
| 11         | 561.5  | 14.6        | 412    | 2 B65134 | protein transport  |
| 12         | 543    | 14.1        | 374    | 2 AC0019 | probable membrane  |
| 13         | 516    | 13.4        | 445    | 2 H64067 | hypothetical prote |
| 14         | 439.5  | 11.4        | 430    | 1 Z4B933 | gene 430 protein - |
| 15         | 370.5  | 9.6         | 660    | 2 B34469 | pullulanase secret |
| 16         | 364.5  | 9.5         | 678    | 2 S46963 | exed protein - Aer |
| 17         | 363    | 9.4         | 1285   | 2 B72420 | hypothetical prote |
| 18         | 346.5  | 9.0         | 649    | 2 S32858 | outD protein - Erw |
| 19         | 337    | 8.8         | 658    | 2 S39553 | xcpQ protein - Pse |
| 20         | 336    | 8.7         | 705    | 2 F70352 | conserved hypothet |
| 21         | 333.5  | 8.7         | 775    | 2 B82471 | general secretory  |
| 22         | 332    | 8.6         | 654    | 2 H65325 | probable general s |
| 23         | 330    | 8.6         | 712    | 2 B47021 | pectic enzyme secr |
| 24         | 329.5  | 8.6         | 591    | 2 S64727 | protein secretion  |
| 25         | 328    | 8.5         | 710    | 2 S28014 | outD protein - Erw |
| 26         | 322.5  | 8.4         | 642    | 2 T42133 | type II secretion  |
| 27         | 319.5  | 8.3         | 687    | 2 D87370 | general secretion  |
| 28         | 318.5  | 8.3         | 625    | 2 E70411 | general secretion  |
| 29         | 311    | 8.1         | 823    | 2 AD1935 | general secretion  |

30 309 8.0 428 2 S08084 gene IV protein -  
31 307.5 8.0 640 2 AF0100 general secretion  
32 305.5 7.9 759 2 C41843 befD protein - Xan  
33 303 7.9 921 2 F71486 probable yopC/gen  
34 300 7.8 776 2 C83411 secretion protein  
35 298.5 7.8 426 1 Z4BPF1 gene IV protein -  
36 296 7.7 559 2 G82326 MSHA biogenesis pr  
37 294 7.6 737 2 I39547 S-protein secretio  
38 292.5 7.6 426 1 Z4BPF1 gene IV protein -  
39 292.5 7.5 426 1 Z4BPF1 gene IV protein -  
40 290 7.5 437 1 Z4BPF1 gene IV protein -  
41 285 7.4 387 2 B72292 hypothetical prote  
42 284.5 7.4 607 2 C40361 virC-region hypoth  
43 278.5 7.2 607 2 T43573 type III secretion  
44 273.5 7.1 748 2 A45243 envelope protein H  
45 271 7.0 672 2 D81746 type III secretion

ALIGNMENTS

RESULT 1  
A81985  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: A81985  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo:; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre:  
Nature 404, 502-506 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491  
A:Reference number: A81775; MUID:2022256; PMID:10761919  
A:Accession: A81985  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-761 <PAR>  
A:Cross-references: GB:AL162753; GB:AL157959; NID:97379120; PIDN:CAB83938.1; PID:g737  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: pilQ; NMA0650

| Query Match           | 95.3%           | Score 3666;  | DB 2;     | Length 761; |
|-----------------------|-----------------|--|-----------|-------------|
| Best Local Similarity | 96.2%           | Pred. No. 4.3e-203;  |           |             |
| Matches 740;          | Conservative 2; | Mismatches 19;   | Indels 8; | Gaps 1;     |
| QY                    | 1               | MNTKLTKIISGLFVATAAFQTASAGNITDIKVSLPNKQKIVKVSFDEKEIYNPTGFTVSS | 60        |             |
| Db                    | 1               | MNTKLTKIISGLFVATAAFQTASAGNITDIKVSLPNKQKIVKVSFDEKEIYNPTGFTVSS | 60        |             |
| QY                    | 61              | PARIALDFEQTGISMDQOVLEYADPLLSKISAAQNSSRARLVNLNKPQYNTVEVRGNKV  | 120       |             |
| Db                    | 61              | PARIALDFEQTGISMDQOVLEYADPLLSKISAAQNSSRARLVNLNKPQYNTVEVRGNKV  | 120       |             |
| QY                    | 121             | WIFINESDDTVSAPARPAVKAAPAKQOQCRVYQVRSIRIQTLYPGKTTAAAPFTES     | 180       |             |
| Db                    | 121             | WIFINESDDTVSAPARPAVKAAPAKQOQCRVYQVRSIRIQTLYPGKTTAAAPFTES     | 180       |             |
| QY                    | 181             | VVSUSAPSPAKQQAASAKQOATAAPAKQOATAAPAKQOATAAPAKQOATAAPAKQOATA  | 240       |             |
| Db                    | 181             | VVSUSAPSPAKQQAASAKQOATAAPAKQOATAAPAKQOATAAPAKQOATAAPAKQOATA  | 240       |             |
| QY                    | 241             | ELALGFAGQPDISQCHDHIIVTLKNHTLPTTLQSLDVADEKTPVQKVTUKRLNNDTQL   | 300       |             |
| Db                    | 241             | ELALGFAGQPDISQCHDHIIVTLKNHTLPTTLQSLDVADEKTPVQKVTUKRLNNDTQL   | 300       |             |
| QY                    | 301             | IITTAGNVELYNKSAAGYFTFQVLPKONLESQVNNAPKFTFGKRLSLDFQDVEIRTI    | 360       |             |
| Db                    | 301             | IITTAGNVELYNKSAAGYFTFQVLPKONLESQVNNAPKFTFGKRLSLDFQDVEIRTI    | 360       |             |
| QY                    | 361             | LOILAKESGMNIVASDSVNGKMTLSLKDVPWDQALDLVQARNLDMRQGNIVNIAPRDE   | 420       |             |
| Db                    | 361             | LOILAKESGMNIVASDSVNGKMTLSLKDVPWDQALDLVQARNLDMRQGNIVNIAPRDE   | 420       |             |

421 LLAKDRAFLQAEDIALDGLALYSQNFQKYNVEEFSILRLDNDADTTGNRTNLVSGRGS 480  
 413 LLAKDRAFLQAEDIALDGLALYSQNFQKYNVEEFSILRLDNDADTTGNRTNLVSGRGS 472  
 481 VLIDPATNTLIVTDRSVIEKFKLIDELDPVPAQVMIEARIVEAADGFSRDLGVKFGAT 540  
 473 VLIDPATNTLIVTDRSVIEKFKLIDELDPVPAQVMIEARIVEAADGFSRDLGVKFGAT 532  
 541 GKXKLKNDTSAFGWGNSGFGGDDKWAETKINLPITAAANSISLVRAISSGALNLELSA 600  
 533 GKXKLKNDTSAFGWGNSGFGGDDKWAETKINLPITAAANSISLVRAISSGALNLELSA 592  
 601 SESLSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSIANGSGSTNTEKKAIVGLTUTVP 660  
 593 SESLSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSIANGSGSTNTEKKAIVGLTUTVP 652  
 661 NITPDGQIIMTVKINKDSPAQCSAGNTILCISTKNLNTQAMVENGGLTIVGGIYEEDNG 720  
 653 NITPDGQIIMTVKINKDSPAQCSAGNTILCISTKNLNTQAMVENGGLTIVGGIYEEDNG 712  
 721 NTLTKVPLLLGDIPIVGNLFKTRGKKTDRRELLIFITPRIMGTAGNSLRY 769  
 713 NTLTKVPLLLGDIPIVGNLFKTRGKKTDRRELLIFITPRIMGTAGNSLRY 761

RESULT 2  
 S70838  
 pilQ protein - Neisseria gonorrhoeae  
 C:Species: Neisseria gonorrhoeae  
 C>Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 08-Oct-1999  
 C:Accession: S70838; S70835  
 R:Drake, S.L.; Koomey, M.  
 submitted to the EMBL Data Library, November 1995  
 A:Description: The product of the pilQ gene is essential for the biogenesis of type IV pili in  
 A:Reference number: S70838  
 A:Accession: S70838  
 A:Molecule type: DNA  
 A:Residues: 1-720 <DRA>  
 A:Cross-references: EMBL:U40596; NID:g1173872; PIDN:AAC43603.1; PID:g1173875  
 R:Drake, S.L.; Koomey, M.  
 Mol. Microbiol. 18, 975-986, 1995  
 A:Title: The product of the pilQ gene is essential for the biogenesis of type IV pili in  
 A:Reference number: S70835; MUID:96422484; PMID:8825101  
 A:Accession: S70835  
 A:Molecule type: DNA  
 A:Residues: 140-214 <DRW>  
 A:Cross-references: EMBL:U40596; NID:g1173872; PIDN:AAC43603.1; PID:g1173875  
 A:Experimental source: strain MS11  
 C:Genetics:  
 A:Gene: pilQ

Query Match 77.7%; Score 2988; DB 2; Length 720;  
 Best Local Similarity 81.7%; Pred. No. 3.9e-164;  
 Matches 619; Conservative 21; Mismatches 78; Indels 40; Gaps 4;

QY 1 MNTKLTIIISGLFVATAAFQTASAGNITDIKVSSLPNKQKIVKVSFDEKEIVNPTGFTSS 60  
 DB 1 MNTKLTIIISGLFVATAAFQTASAGNITDIKVSSLPNKQKIVKVSFDEKEIVNPTGFTSS 60  
 QY 61 PARIALDFEQTGISMDQVLEYADPILLSKISAAQNSRRARLVNLNKPQGYNTEVRGNKV 120  
 DB 61 PARIALDFEQTGISMDQVLEYADPILLSKISAAQNSRRARLVNLNKPQGYNTEVRGNKV 120  
 QY 121 WIFINESDDTVSAPAPPAKPAKAAQKQGRVYQVRSIRIQTLYPGKTTAAAPFTES 180  
 DB 121 WIFINESDDTVSAPAPPAKPAKAAQKQGRVYQVRSIRIQTLYPGKTTAAAPFTES 158  
 QY 181 VVSVSAPSPAKQQAASAKQQTAAAPAKQQAAPAKQQTATNIDFRKDGKNAGII 240  
 DB 159 VVSVSAPSPAKQQAASAKQQAASAKQQTAAAPAKQQAAPAKQQTATNIDFRKDGKNAGII 202  
 QY 241 ELAALGFAGOPDISQOHDHIIIVTLKNHTLPTTLQRSLDVADFKTPVQKVTLKLNNDTOL 300

203 ELAALGFAGOPDISQOHDHIIIVTLKNHTLPTTLQRSLDVADFKTPVQKVTLKLNNDTOL 262  
 301 IITTAGNWLKNSAAGPYFTFQVLPKKQNLSEGGVYNNAPKFTTGRKISLDQFQVEIRTI 360  
 263 IITTAGNWLKNSAAGPYFTFQVLPKKQNLSEGGVYNNAPKFTTGRKISLDQFQVEIRTI 322  
 361 LQILAKESGNNIVASDSVNGKMTLSLKDVPWDQALDLVMOQARNLDMRQOQGNIVNAPRDE 420  
 323 LQILAKESGNNIVASDSVNGKMTLSLKDVPWDQALDLVMOQARNLDMRQOQGNIVNAPR-R 381  
 421 LLAKDRAFLQAEDIALDGLALYSQNFQKYNVEEFSILRLDNDADTTGNRTNLVSGRGS 480  
 382 AACQKSLTSGKXHCGRSRAVFPKLPKIEIQKGRIPQHPALDNDADTTGNRTNLVSGRGS 441  
 481 VLIDPATNTLIVTDRSVIEKFKLIDELDPVPAQVMIEARIVEAADGFSRDLGVKFGAT 540  
 442 VLIDPATNTLIVTDRSVIEKFKLIDELDPVPAQVMIEARIVEAADGFSRDLGVKFGAT 501  
 541 GKXKLKNDTSAFGWGNSGFGGDDKWAETKINLPITAAANSISLVRAISSGALNLELSA 600  
 502 GRKXKLKNETSAFGWNSGFGGDDKWAETKINLPITAAANSISLVRAISSGALNLELSA 560  
 601 SESLSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSIANGSGSTNTEKKAIVGLTUTVP 660  
 561 SESLSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSIANGSGSTNTEKKAIVGLTUTVA 620  
 661 NITPDGQIIMTVKINKDSPAQCSAGNTILCISTKNLNTQAMVENGGLTIVGGIYEEDNG 720  
 621 NITPDGQIIMTVKINKDSPAQCSAGNTILCISTKNLNTQAMVENGGLTIVGGIYEENNG 680  
 721 NTLTKVPLLLGDIPIVGNLFKTRGKKTDRRELLIFITPR 758  
 681 NTLTKVPLLLGDIPIVGNLFKTRGKKTDRRELLIFITPR 718

RESULT 3  
 A37051  
 outer membrane protein-molecular complex (omc) precursor - Neisseria gonorrhoeae  
 C:Species: Neisseria gonorrhoeae  
 C>Date: 31-Jan-1992 #sequence\_revision 31-Jan-1992 #text\_change 17-Feb-1994  
 C:Accession: A37051  
 R:Tsal, W.M.; Larsen, S.H.; Wilde III, C.E.  
 Infect. Immun. 57, 2653-2659, 1989  
 A:Title: Cloning and DNA sequence of the omc gene encoding the outer membrane protein  
 A:Reference number: A37051; MUID:89339707; PMID:2503445  
 A:Accession: A37051  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-711 <TSA>  
 A:Cross-references: GB:L19944; GB:M22564  
 C:Keywords: membrane protein

Query Match 71.9%; Score 2765.5; DB 2; Length 711;  
 Best Local Similarity 76.0%; Pred. No. 2.4e-151;  
 Matches 576; Conservative 29; Mismatches 104; Indels 49; Gaps 5;

QY 1 MNTKLTIIISGLFVATAAFQTASAGNITDIKVSSLPNKQKIVKVSFDEKEIVNPTGFTSS 60  
 DB 1 MNTKLTIIISGLFVATAAFQTASAGNITDIKVSSLPNKQKIVKVSFDEKEIVNPTGFTSS 60  
 QY 61 PARIALDFEQTGISMDQVLEYADPILLSKISAAQNSRRARLVNLNKPQGYNTEVRGNKV 120  
 DB 61 PARIALDFEQTGISMDQVLEYADPILLSKISAAQNSRRARLVNLNKPQGYNTEVRGNKV 120  
 QY 121 WIFINESDDTVSAPAPPAKPAKAAQKQGRVYQVRSIRIQTLYPGKTTAAAPFTES 180  
 DB 121 WIFINESDDTVSAPAPPAKPAKAAQKQGRVYQVRSIRIQTLYPGKTTAAAPFTES 172  
 QY 181 VVSVSAPSPAKQQAASAKQQTAAAPAKQQAAPAKQQTATNIDFRKDGKNAGII 240  
 DB 173 VVSVSAPSPAKQQAASAKQQTAAAPAKQQAAPAKQQTATNIDFRKDGKNAGII 194  
 QY 241 ELAALGFAGOPDISQOHDHIIIVTLKNHTLPTTLQRSLDVADFKTPVQKVTLKLNNDTOL 300



```
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
M.; Tsuchida, M.H.; Vailada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0373

Query Match 23.6%; Score 907; DB 2; Length 637;
Best Local Similarity 34.5%; Pred. No. 1.5e-44;
Matches 203; Conservative 127; Mismatches 221; Indels 38; Gaps 11;

QY 207 AKQQTAAAPAKQAAAPAKQTNIDPRKQGNAGIIELAAAGFAGQPDIDSQQHDHIIIVTKN 266
Db 48 AENKQAIIPVKVANNAPLSVSKIDFKGGDGGRLIKFKDGGATPDLATQGTTLVLDLT 107
QY 267 HTLPTLQSLDVAFTKPVQKVTUKLNNDTQLIITTAGHWELVNSAAPFFQVLP 326
Db 108 AVLPTVLQRIINVDVATPVQIDAKPMGKQALVSTKGAFESLAYQTGDEY-VWEIVP 166
QY 327 KQNLESQGV-----NNAPKT---FTGRKISLDFODVEIRTIQILAKESGNIVAS 375
Db 167 RKGEAAAGGAIPTESVSKAAKISAGYSGRPVTFNQDPVPTVTLQIADSNLIVAS 226
QY 376 DSVNGKWTLSLKDVPWDQALDLMQARNLDMQOQGNVNIAPRDELLAKDKAFLOAKDI 435
Db 227 DTQGNVTLLINVPWDQALDIVLRAKGLDKRDEKVIWVAPQQLAKYEQEKEDARIAI 286
QY 436 ADGALYSONFQLYKYNVEE-FRSLRLDNADTTGN-----RNTLVSGRSLVIDPAT 487
Db 287 ENREGIITVQINYHSATVIFKALTAKMGGGGNGTNNNSDDAFSLSPRGELVADERT 346
QY 488 NTLIVDTSRVIEKFKRLIDELDPAAQVMIEARIVEAADGFSRDLGVKFGATGKKKLN 547
Db 347 NTLMSIDIPKVAQMTLIQHIDRPVDQVLIIEGRIVIAITDSFARDLGAKFGVGAASRFS 406
QY 548 DTSFAGVNGSGGGDDKWAETKINLPITAAANSISLVRAISSGA-----LNLEL 598
Db 407 NTATIGSHVTT---ADSSSTATKGLNVDLGGPTNTTASV--LPSLAYTLGPKFNLDLEL 461
QY 599 SASESLSKTKLANPVLTONRKEAKIESGYEIPF-TVT-SIANGSSSTNTELKKAVALGL 656
Db 462 SALQOESRGEVINPRIVANQREGYIKQKEIGYVITIGGVAGGQATPNVQKEAVLEL 521
QY 657 TVTPNITPDGQIIMTVKINKDSPAQ--CASGNQITLICISTNLTQAMVENGCTLIIVGGI 714
Db 522 RVTPITINDNRVFLNMTVKKDEIDQMITTANFGTVPLNKREINTAVLVDDGQTVVIGV 581
QY 715 YEEDNGNTLTKVPLLDGIPVIGNLFTKRGKTDRELLIFITPRIMGTA 763
Db 582 YEFSDRNSVSKVFLGDVFPFLGNLFPKRGDKQKAKELLIFVTPKVLRA 630

RESULT 7
D82816
fimbrial assembly protein VC2630 [imported] - Vibrio cholerae (strain N16961 serogro
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: D82053
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
1. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: D82053
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-578 <HEI>
A:Cross-references: GB:AE004329; GB:AE003852; NID:9657211; PIDN:AAF95771.1; GSPDB:G
```

A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC2630  
A:Map position: 1

Query Match 21.0%; Score 809.5; DB 2; Length 578;  
Best Local Similarity 34.3%; Pred. No. 5.3e-39;  
Matches 191; Conservative 128; Mismatches 205; Indels 33; Gaps 14;  
QY 218 QAAAPAKOTNIDFRKDGKNAGIIEAALGFGAGOPDISOHOHIIIVTLKXHTLPTTLQSL 277  
DB 39 ESATANQLENIDFRVKEKAIVLVELASPSAVDVQKVQEGSLTELLKTDVADKLYLL 98  
QY 278 DVADFKTPQVKYTLKRLNNDTOLIITTAGNMELVNKSAAPGYFTQVLPKQNLBSGGVN 337  
DB 99 DVKDFSTPVSVEFRKEPSTQLVTVTDG--BFQHDYTLKGYLEWISKLKADP---K 152  
QY 338 NAKPTF---TGRKISLDFODVEIRTIQILAKESGMNIVASDSVNGKMTLSLKVPMDOA 394  
DB 153 PKPSVLEKEGLISINFQDIIPRVNVLQIADYNGFNLVSDVVGNTLRLDGVPMQOV 212  
QY 395 LDLVQARNLDRQOGNIVNIAPRDELLAKOKAFLOAKDIAADGALYSQNFQKLYKNWE 454  
DB 213 LDIIQVKGLDKRVGDNVLIAPKEELDREKQALEKARLAEGLDLKSEIIKINFARAS 272  
QY 455 EFRSILRLDNADTTGNRTLVSGRSLVDPATNTLIVTDRSVIEKFKRLIDELVPAQ 514  
DB 273 DIAAMI-----GEGGNV-NLSESGISIDERTNLSLLRELDPNTAVIREIIESLDIPVK 326  
QY 515 QVMIEARIVEAADGFSRDLGVKFGATKKKLNKNTSAFGWGNV-----GFGDDKWQAE 569  
DB 327 QVOIEARIIVTKEGNLEELGVWGVMS---TNGSHSVGGSTESNLWQKGLLADDEFPVD 382  
QY 570 ---TKINLPITAA-ANSISL-VRAISSGA-LNLELSASELSKTKTLANPRVLVTQNRKEAK 624  
DB 383 EFLNVNLASTSANASSIAFQVAKLGGTLDLELSALQESKAEIISPRLLITNKQPAY 442  
QY 625 IESGYEIPFTVTSIANGSGSTNTLKAVALGLTVPNTIPDQGIIMTVKINKDSPAQCA- 683  
DB 443 IEQGTIIPYLESS--SSGAST-VAPKAVLSLKVTPQITPDNRVLVLSVTDQRRGETVK 499  
QY 684 SGNQTLICSTKNLTQAMVNGGTLVGVIEEDNGNTLTKVPLLGDIPIVGNLFKRTG 743  
DB 500 TGTGEAVSIDTQIGQVLVNGETVVLGGIFQHSINNNSVDKVPILGDLPLVIGALFRTY 559  
QY 744 KKTDRRELLIFITPRIM 760  
DB 560 EQMGKSELLIFVTPKV 576

RESULT 8  
AB1000  
Type II secretion system protein [imported] - Salmonella enterica subsp. enterica serovar Typhimurium  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AB1000  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, J.; Moul, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AB1000  
A:Molecule type: DNA  
A:Status: preliminary  
A:Residues: 1-412 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD08126.1; PID:g16505105; GSPDB:GNO0176  
C:Genetics:  
C:Superfamily: hypothetical protein HI0435

Query Match 14.9%; Score 574.5; DB 2; Length 412;  
Best Local Similarity 32.8%; Pred. No. 1e-25;  
Matches 139; Conservative 94; Mismatches 148; Indels 43; Gaps 10;

QY 347 KISLDFODVEIRTIQILAKESGMNIVASDSVNGKMTLSLKVPMDOALDVLVMOARNLDM 406  
DB 23 KVTLVDDVPVQVLOALAEQERQNLVSPDVSGTSLSLHLDVPMQALQTVVNSAGLV 82  
QY 407 RQOGNIVNIA-----PRDELLAKOKAFLOAKDIAADGALYSQNFQKLYKNVEFRSILRL 462  
DB 83 RQEGNILVHVSQAWKEHSARQDAERLRLOANL-----PLENRSISIQYADAGELAK---- 134  
QY 463 DNADTTGNRTLVSGRSLVDPATNTLIVTDRSVIEKFKRLIDELVPAQVMIEARI 522  
DB 135 -----AGEK--LLSAGKTINVKKTNRLLDRNRAALAELEKWKVSMQMDLPVAVELAHI 187  
QY 523 VEADGFSRDLGVKFGATKKKLNKNTSAFGWGNVSGFGDDKQWGAETKINLITAAANS 582  
DB 188 VTINEKSLRELGVKWK-----TLADATQAGAVGDTVTLSSD-----LSVAAATSR 231  
QY 583 ISL-VRAISSGALNLELSASELSKTKTLANPRVLVTQNRKEAKIESGYEIPFTVTSIANG 641  
DB 232 VGNFNGRNLRLDLELSALEQKQOLDIIASPRLLASHLPASIKQSGSEIPYQVSSGESG 291  
QY 642 GSSTNTLKAVALGLTVPNTIPDQGIIMTVKINKDSPAQCA---ASGNQTLICISTKNLN 698  
DB 292 --ATSVFEKAVLGMETP-VLQKGRIRLKLHISQNVPGQVLQOADGE--VLAIDKQEI 347  
QY 699 TQAMVNGGTLVGVIEEDNGNTLTKVPLLGDIPIVGNLFKRTGKKTDRRELLIFITPR 758  
DB 348 TQVEVKSGETLALGGIFSRKNKSGSDSVPLLGDIPIWGLQFLRHDKEDERELVVFITPR 407  
QY 759 IMGT 762  
DB 408 LVAT 411

RESULT 9  
A91158  
F:obable transport portein BCs4233 [imported] - Escherichia coli (strain O157:H7, subsp. C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C. gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: A91158  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-412 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA937656.1; PID:g13363707; GSPDB:GNO0154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: ECs4233  
C:Superfamily: hypothetical protein HI0435

Query Match 14.7%; Score 567.5; DB 2; Length 412;  
Best Local Similarity 33.5%; Pred. No. 2.6e-25;  
Matches 142; Conservative 94; Mismatches 147; Indels 41; Gaps 11;  
QY 346 EKISLDFODVEIRTIQILAKESGMNIVASDSVNGKMTLSLKVPMDOALDVLVMOARNL 405  
DB 22 QKVTLVDDVPVQVLOALAEQERQNLVSPDVSGTSLSLHLDVPMQALQTVVNSAGLI 81  
QY 406 MRQOGNIV---NIAPRDELLAKOKAFLOAKDIAADGALYSQNFQKLYKNVEFRSILRL 462  
DB 82 TRQEGNILSVHSIAQWQNDNIARQEA--EQARAQANL-PLENRTITLOVADAGELAK---- 134  
QY 463 DNADTTGNRTLVSGRSLVDPATNTLIVTDRSVIEKFKRLIDELVPAQVMIEARI 522  
DB 135 -----AGEK--LLSAGKSNVTVDKTRNRLLRDNKLTALSALQWVAQMDLPVQVLSAHI 187

Db 348 TQVEKSGETLALGGIFTRKNKGSDSVPLGLDIPWFGQLFRHDGKEDERELVVFITPR 407

QY 759 IMGT 762  
:::  
Db 408 LVSS 411

RESULT 11  
B65134  
protein transport protein hofq precursor - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C>Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: B65134  
R;Blattner, F.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
A.: Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A>Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:3278503  
A:Accession: B65134  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-412 <BLAT>  
A:Cross-references: GB:AEO00414; GB:U00096; NID:G1789783; PIDN:AAC76416.1; PID:G1789783  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: hofQ  
C:Superfamily: hypothetical protein HI0435

Query Match 14.6%; Score 561.5; DB 2; Length 412;  
Best Local Similarity 32.7%; Pred. No. 5.8e-25;  
Matches 141; Conservative 89; Mismatches 146; Indels 55; Gaps 10;

QY 346 RKISLDPOVEIRTIQILAKESGMNIVASDSVNGKMTLSLKDVDPWDQALDLVMQARNLD 405  
:::  
Db 22 QKVTLMDVDVPVAQVLQALAEQEKLVVSPDVGTVSLHLTDVPMQALQTVMKSAGLI 81

QY 406 MRQQGNIV-----NIAPDELLAKDKAFLOAEKDIALDGALYSQNFKLYNNVEE 455  
:::  
Db 82 TRQEGNILSVHSIAWNQNNIARQEAQAQANLPLE-----NRSITLOYADAGE 131

QY 456 FRSTRLDNADTTGNRTNLVSGRGSVLIDPATNTLTIVTDRSVEIKFRLIDELDVPAQO 515  
:::  
Db 132 LAK-----AGEK--LLSAKGMTVKTRNLRLLRNKTALSALQEWQAMDLPVQG 180

QY 516 VMIEARIVEAADGFSDLGKFGATGKKLNDTSFAFGVNSGFGDDKWGAETKINLP 575  
:::  
Db 191 VELSAHIVTNEKSLRELGVKM-----TLADAQHAGGVQVVTLLGSD-----LS 224

QY 576 ITAAANSISL-VRAISSGAINLSLSSESLSKTKTLANPRVLTQNRKEAKIESGYEIPT 634  
:::  
Db 225 VATATTTHVGFINGRLLDLSALEEQXQQLDIITASPRLLASHLQPASIKOGSEIPIQ 284

QY 635 VTSTANGSSNTTELKAVLGLTVTPNITPDGOIMTVKINKDSPAQ-----ASNQOTILC 691  
:::  
Db 285 VSSGESG--ATSVEFEKAVLGMETPVLQKGRILKLHISQNVPGVQLQADGE--VLA 340

QY 692 ISTKNLNTQAMVENGTLIVGGIYEENGNLTLPULLGDIPVIGLFTKRGKKTDBREL 751  
:::  
Db 341 IDKQBIETQVEKSGSETLALGGIFTRKNKGSDSVPLGLDIPWFGQLFRHDGKEDEREL 400

QY 752 LIPIITPRIMG 762  
:::  
Db 401 VVFITPRUVSS 411

RESULT 12  
AC0019  
probable membrane transport protein YPO0150 [imported] - Versinia pestis (strain CO92)  
C:Species: Versinia pestis  
C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
C:Accession: AC0019  
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.;  
deno-Farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan,



il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001  
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A:Reference number: AB0001; MUID:21470413; PMID:11586360  
 A:Accession: AC0019  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-374 <KUR>  
 A:Cross-references: GB:AL590842; PIDN:CAC89013.1; PID:g15978255; GSPDB:GN00175  
 A:Genetics:  
 A:Gene: YPO0150  
 C:Superfamily: hypothetical protein HI0435

Query Match 14.1%; Score 543; DB 2; Length 374;  
 Best Local Similarity 31.6%; Pred. No. 5.8e-24;  
 Matches 127; Conservative 91; Mismatches 152; Indels 32; Gaps 7;

QY 361 LQILAKESGMNIVASDSVNGKMTLSKQVDPDQALDLVQWQARNLDMRQGNIVNAPRDE 420  
 DB 1 MQALADYRQLNLTITGVGGNLSRLIEVPWEQALAILRMGRKAEREIVMMVFTTQE 60  
 QY 421 LLAQKAFQAEKDIADLQALYSQNFQKQVVEFRGILRDNDADTTGNRNLTVSGRGS 480  
 DB 61 IQERQ----QRTKQAAPALANLTLALQYANAQV----ADSLDPL--QGLLSPLGS 109  
 QY 481 VLIDPATNTLIVTDRSVIEKFRKILDELDPVPAQVMIEARIVEAADGFSRDLGVKFGAT 540  
 DB 110 VVADKRTNLTLLRDPASLALKLMLIENDLPQQVQLSAHTVITSSDLOELGVR---- 165  
 QY 541 GKKLKNDIFSAGWGVNSGFGDDKQWAEKTKINLPITAAANSISL-VRAISSGALNLELS 599  
 DB 166 -----WNGEGKGNTALRINDFNVLPLPNSAASVGFHVARIGRLLLELS 212  
 QY 600 ASBSLSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSIANGSGSTNTELKAVJGLVT 659  
 DB 213 ALEQENQVDIIASPRUITSHQQTASIKQSDIPYTVS--RGKEAAAEFKEAVLGMVET 270  
 QY 660 BNITPDGQIIMTVKINKDSPA-QCASGNQITLCISTKNLNTQAMVNGGTLVGGIYEBD 718  
 DB 271 PKLRNGKILDLKISQWPGITIKGESEMLLDKQEIKTQTVNDGETIVLGGIFQOK 330  
 QY 719 NGNTLTKVPLGDPVIGMLFKTRGKTKDRRELLIIPTRIM 760  
 DB 331 KRQSVNKKVPLADIPULGAMFRQDTQQQSRRELIVITPKLI 372

RESULT 13  
 H64067  
 hypothetical protein HI0435 (transformation locus) - Haemophilus influenzae (strain Rd K  
 N:Alternate names: ORFE protein  
 C:Species: Haemophilus influenzae  
 C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 29-Sep-1999  
 C:Accession: H64067; JH0434  
 C:Authors: J.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, J.  
 ; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
 ; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
 Science 269, 496-512, 1995  
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
 A:Reference number: A64000; MUID:95350630; PMID:7542800  
 A:Accession: H64067  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-445 <TIGR>  
 A:Cross-references: GB:U32726; GB:L42023; NID:g1573399; PIDN:AA22094.1; PID:g1573410; T  
 A:Experimental source: strain Rd KW20  
 R:Tomb, J.P.; El-Hajj, H.; Smith, H.O.  
 Gene 104, 1-10, 1991  
 A:Title: Nucleotide sequence of a cluster of genes involved in the transformation of Ha  
 A:Reference number: JH0430; MUID:92009183; PMID:1916268  
 A:Accession: JH0434  
 A:Molecule type: DNA  
 A:Residues: 1-432, 'NVRGVET', 440, 'K', <TOM>

A:Cross-references: GB:M62809; NID:g148990; PIDN:AAA5012.1; PID:g148997  
 A:Experimental source: strain Rd  
 C:Superfamily: hypothetical protein HI0435

Query Match 13.4%; Score 516; DB 2; Length 445;  
 Best Local Similarity 29.5%; Pred. No. 2.7e-22;  
 Matches 132; Conservative 93; Mismatches 165; Indels 58; Gaps 12;

QY 340 PKFTGRKISLDFQDVEIRTIQILAKESGMNIVASDSVNGKMTLSKQVDPDQALDLV 399  
 DB 25 PKT-DNERFFIRLSQAPLAQTLQSLAQDQVNLVIGILENKISLKNINIMPRLLQIIA 83  
 QY 400 QARNLDMRQ-----GNIVNAPRDELLAKDKAFQAEKDIADLQALY 442  
 DB 84 KSKHLTJNKDDGYLYNGSOGKQVAGNLTNTEPH-----LV 121  
 QY 443 SQNFQKQVVEFRSILRDNDADTTGNRNLTVSGRSLVLDIPATNTLIVTDRSVIEKF 502  
 DB 122 SHTVKLHFAKASLMKSL-----TTGS-GSLLSPAGSITFDRSNLLVTDQSPRSQNI 174  
 QY 503 KKLDELDPVPAQVMIEARIVEAADGFSRDLGVKFGATGKKLKNDSARSAGWGVNSGFG 562  
 DB 175 KKLIAEMDKPIEQIAIEARIVITDESLEKELGVWVGIF--NPTENARRVAGSUTGNSF-- 230  
 QY 563 DDKQWAEKTKINLPITAA-ANSISL-VRAISSGALNLELSASELSKTKTLANPRVLTONR 620  
 DB 231 -ENIADNLNVNFATTTTPAGSIALQVAKINGRLDLLELSALERENNVETIASPRLLTTNK 289  
 QY 621 KEAKIESGYEIPFTVTSIANGSGSTNTELKAVJGLVTNITPDGQIIMTVKINKDSPA 679  
 DB 290 KSASIKQGTIPYTVSNTN--DTQSVFEFRAVLGLEVTTHISKDNILLDLVLSQSPG 347  
 QY 680 AQCASGNQITLCISTKNLNTQAMVNGGTLVGGIYEBDNGNTLTKVPLGDPVIGMLF 739  
 DB 348 SRVAYGQNEVSDIKQEIQTQVFAKQGETIVLGGVFDHTITKSEDKVPLGDPVIGMLF 407  
 QY 740 KTRGKTKDRRELLIIPTRIMGTAGNSL 767  
 DB 408 SKESERHQKRELIVFVTPHIL-KAGETL 434

RESULT 14  
 Z4BP33  
 gene 430 protein - phage Pf3  
 C:Species: phage Pf3  
 A:Note: host Pseudomonas aeruginosa  
 C:Date: 28-May-1986 #sequence\_revision 28-May-1986 #text\_change 24-Sep-1999  
 C:Accession: A04270  
 R:Luitel, R.G.M.; Putterman, D.G.; Schoenmakers, J.G.G.; Konings, R.N.H.; Day, L.A.  
 J. Virol. 56, 268-276, 1985  
 A:Title: Nucleotide sequence of the genome of Pf3, an IncP-1 plasmid-specific filament  
 A:Reference number: A94693; MUID:85293231; PMID:3928901  
 A:Accession: A04270  
 A:Molecule type: DNA  
 A:Residues: 1-430 <LUI>  
 A:Cross-references: GB:M11912; NID:g215371; PIDN:AAA89381.1; PID:g215375  
 C:Comment: Bacteriophage Pf3 is a class II filamentous phage.  
 C:Genetics:  
 A:Gene: 430  
 A:Superfamily: filamentous phage gene IV protein  
 C:Keywords: phage maturation

Query Match 11.4%; Score 439.5; DB 1; Length 430;  
 Best Local Similarity 27.5%; Pred. No. 6.4e-18;  
 Matches 118; Conservative 93; Mismatches 175; Indels 43; Gaps 10;

QY 340 PKFTGRKISLDFQDVEIRTIQILAKESGMNIVASDSVNGKMTLSKQVDPDQALDLV 399  
 DB 16 PFAFASDRLTVKHHEIDIRVAIPVADFCGRSVLGPISIQGVSLDDPDDVPCSQAFDILL 75  
 QY 400 QARNLDMRQGNIVNAPRDELLAKDKA---FLOAEKDIADLQALYSQNFQKQVVEFR 456

Db 76 ESNHLLSSMGVDVLVITAMQVLNSERKADDLRTFRDLFNANDIERRVINIVHASASEV 135  
Qy 457 RSLR--LDNADTTGNRTNLVSGRSVLDPATNTLIIVTDRSVIEKPKLIDELVPAQ 514  
Db 136 VSLFKESFMSLDAPG-----SMTVDERTNSVFAALDSSFPALLESVIQALDVPR 186  
Qy 515 QYMIARIIEAADGFSRDLGVKFGATGKKLKNDSAFSGWVNSFGGDKWGAETKINL 574  
Db 187 QVAIEANVVEASVDKSLGLNAG--GALSNG-----WSAVTAGDL 226  
Qy 575 PTAAANSISLVRAISSGALNLE--LSASESLKTKLANPRVLTONRKEAKIESGYEIP 632  
Db 227 SV-AAGSSIGF--GFLSNTLSLDGFTAMENGNRGRVSRPFLLLTLDKQASVLRGTLP 283  
Qy 633 FTVTSIANGSSSTNELKXKVLGLTVTPNITPDGOIIMTKVINKDSPACASGNQITLCI 692  
Db 284 YQOSA---GGGATSVAFKHAALSLEVKVPSDINSIVIEVLVSRDSP-NFSNAIDGVPEI 339  
Qy 693 STKXNLTQAWENGGLIIVGGIYEEDNGNTLTKVPLGDIPIVIGLNFTRGKKTDERELL 752  
Db 340 DTNRLVTIRVRPHGQTVLGGVYSTINQOGSSRSVGSIRIPGIGRLFKKHEHTQYELL 399  
Qy 753 IFITPRIMG 761  
Db 400 IFITPRILG 408  
RESULT 15  
B34469  
Pullulanase secretion protein puld precursor - Klebsiella pneumoniae (strain UNF5023)  
C:Species: Klebsiella pneumoniae  
C:Date: 15-Jun-1990 #sequence\_revision 31-Dec-1993 #text\_change 20-Feb-1995  
C:Accession: B34469; B31394  
R: d'Enfert, C.; Reys, I.; Wandersman, C.; Pugley, A. P.  
J. Biol. Chem. 264, 17462-17468, 1989  
A:Title: Protein secretion by gram-negative bacteria. Characterization of two membrane  
A:Reference number: A34469; MUID:30008916; PMID:2677007  
A:Accession: B34469  
A:Molecule type: DNA  
A:Residues: 1-660 <DEN>  
A:Cross-references: GB:M32613  
C:Genetics:  
A:Gene: puld  
A:Start codon: TTG  
Query Match 9.6%; Score 370.5; DB 2; Length 660;  
Best Local Similarity 22.1%; Pred. No. 1.1e-13;  
Matches 152; Conservative 116; Mismatches 261; Indels 159; Gaps 23;  
Qy 110 QYNTVEGRNKVWIFINSDDTVSAPAPKAPAKQCGRTTVQVRSIRIQITLYPG 169  
Db 29 EFSASFKGTIDQEFIN---TVSKNLAKTVIIDPSV---RGITVRSYDMLNEEQYQF 80  
Qy 170 KTTAAAPTESVSVSAPFSPAKQQAASAKQOATAAPAKQOATAAP--AKQQAAPAKQTN 227  
Db 81 FLSVLDTVGFATNN---NNGVLKVRSKDAKTAAPVPVADAPGIGDEWTVRVPLTN 136  
Qy 228 ID-----FRKDGKNAGIIELA-----ALGFAGQPDISOQHIIIVTLKNH-----T 268  
Db 137 VAARDLAPLLQLQNDNAGVGVHVEFSNVLLMTGRAAVIKRLLITIVERYDNDAGDSRVT 196  
Qy 269 LPTTLQRSUDVADFKTPQKVTKLKLNNDTQIIITAGNWLNVKSAAPGVFTQVLPKK 328  
Db 197 VPLSWASADV-----VRLVT--ELNKDT-----SKSALPG----- 225  
Qy 329 QNLESGGVNNAKPTFTGRKISLDFQDVVEIRTIQLAKESGMNIVASDVNGKWTLSLKD 388  
Db 226 -----SMVANVADERTNAVLSVEGEPNRQR----- 251  
Qy 389 VPWDQALDLVQARNLDMRQ--QGN-----IVNIAPRDELLAKDAFLQAEKDIALDG 439  
Db 252 -----IIAMIKOLDRQQAQTQGNTRKVIYLYKAKASDLVEVLGTISSTMQSEKQA----- 300

Qy 440 ALYSQNPQLKYKNVEFRSILRLDNADTTGNRTNLVSGRSVLDPATNTLIIVTDRSVI 499  
Db 301 -----KPVAAALD-----KNIIIAHQG-----THALIIVTAAPDVM 330  
Qy 500 EKFKRLIDELVPAQOVMIEARIEAAGFSRDLGVKFG--ATJKKLKNDSATFAGWGVN 557  
Db 331 NDLERVIAQIDIRPQVLVEAIIAEVQDADGLNGICWANKNAGMTQFTNS-----GLPIS 386  
Qy 558 SGFGDDKWAETKINIPITAAANSISLVRA-ISSGALNLELSASESLKTKTKLANPRVL 616  
Db 387 TATAGANQYNKDGTVSSLSASLSFNGIAAGFYQGNWAMLLTALSSTKNDILATPSIV 446  
Qy 617 TQNRKEAKIESGYEIP--FTVTSIANGSSSTNELKXAV-LGLTVTPNITPDGOIIMTKVI 674  
Db 447 TLDNMEATFNVGQEVPLVLTGTSQTTSGDNIENFTVERKTVGIKLVKPKQINEGSDSVLLBIEQ 506  
Qy 675 NKDSPACASGNQITL--CISTKXNLTQAWENGGLIIVGGIYEEDNGNTLTKVPLGDI 732  
Db 507 EVSSVADAASSTSDLCATENTRTVNAVILVGSGETVVGGLLDKSVSDTADKVPVLLGDI 566  
Qy 733 PVIGNLFTKRGKTDRELLIFITPRIM 760  
Db 567 PVIGALFRSTSKVKSKRNLMLEIFRPTVI 594  
RESULT 16  
S46963  
exed protein - Aeromonas salmonicida  
C:Species: Aeromonas salmonicida  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 08-Oct-1999  
C:Accession: I39678; S46963  
R: Karlyshev, A. V.; MacIntyre, S.  
Gene 158, 77-82, 1995  
A:Title: Cloning and study of the genetic organization of the exe gene cluster of Aer.  
A:Reference number: I39675; MUID:95309729; PMID:7789814  
A:Accession: I39678  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-678 <RES>  
A:Cross-references: EMBL:X80505; NID:9516349; PID:CAA56668.1; PID:9516351  
C:Genetics:  
A:Gene: exed  
Query Match 9.5%; Score 364.5; DB 2; Length 678;  
Best Local Similarity 24.7%; Pred. No. 2.6e-13;  
Matches 124; Conservative 85; Mismatches 170; Indels 124; Gaps 17;  
Qy 353 QDVEIRTIQLIAK---ESGMNIVASDSVN---GKMTLSLK-----DVPWDOAL 395  
Db 133 RNVSVRELAPLLRLQNDNAGGNNVHYDPSNVLLITGRAAVVNRVLVEVVRVYKAGDQEV 192  
Qy 396 DLV-----MQARNLDMRQQGNVNIAPRDELLAKKKA----- 427  
Db 193 DIILKLYASAGEVMVRLVTNLNKGNTQGGNTSLLLPKVVADERTNSVVVSGEPKARARI 252  
Qy 428 ---FLQAEKDIALGALYSONFOLKYKN---VEEFRSILRLDNADTTGNRTNLVSGRS 480  
Db 253 IQMTRQLDRDLQSQG--NTRVEFLKYKRAKDMVEVLKGVSTSEADKGGGTTAGGNGAS 310  
Qy 481 -----VLIDPATNTLIIVTDRSVIEKPKLIDELVPAQOVMIEARIEAAGFSRDL 533  
Db 311 IGGGKLAIASDETTNALVITAQPDYMAELEQVAVKLDIRRAQVLEAIIVEIADGDGLNL 370  
Qy 534 GYKFGATGKKLKNDSAFSGWVNSFGGDDKXWGAETKINLPI-----T 577  
Db 371 GQWANTN-----GGGTQF--TDTNLPIGSVAIAAKDYNGTTT 408  
Qy 578 AAANSISLVRAISSG-----ALNLELSASESLKTKTKLANPRVLTONRKEAKIESGYE 630  
Db 409 GLADLAKGFNGMAAGFYHGNWAALVTALST---TKSDILSTFSIVTMDNKEASFNVGOE 465  
Qy 631 IPTVTSIANGSSSTNT-----ELKXAVLGLTVTPNITPDGOIIMTKVINKDSPACQ- 681

Db 466 VP-----VQSGSQSTTSQDVFNTERKTGTGKLTVTTPQINEGSDVLLNIEQEVSSVAQK 520  
 Qy 682 CASGNQTI-LCISTKNLNTQAMVNGOTLIVGGIYEDNGNTLTQVPLGIDPVGNIPLFK 740  
 Db 521 QATGATADLGPFTDRTIKNAVLKSGTETVGLGLMDQETQKVKSPVLLGIDPVLGYLFR 580  
 Qy 741 TRGKKTDRRELLIFITPRIMGTA 763  
 Db 581 STNNTSKRNLMVFIRPTILRDA 603

RESULT 17  
 S32858  
 outd protein - Erwinia carotovora  
 C:Species: Erwinia carotovora  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 08-Oct-1999  
 C:Accession: S32858; S31747  
 R:Reeves, P.J.; Whitcombe, D.; Wharam, S.; Gibson, M.; Allison, G.; Bunce, N.; Barallion, M. Microbiol. 8, 443-456, 1993

hypotheical protein TM0088 - Thermotoga maritima (strain MSB8)  
 C:Species: Thermotoga maritima  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
 C:Accession: B72420  
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garret, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.

Nature 399, 323-329, 1999  
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
 A:Reference number: A72200; PMID:99287316; PMID:10360571  
 A:Accession: B72420  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1285 <ARN>  
 A:Cross-references: GB:AS001695; GB:AE000512; NID:G4980569; PIDN:AAD35182.1; PID:G498057  
 A:Experimental source: strain MSB8  
 C:Genetics:  
 A:Gene: TM0088

Query Match 9.4%; Score 363; DB 2; Length 1285;  
 Best Local Similarity 25.0%; Pred. No. 8.2e-13;  
 Matches 105; Conservative 95; Mismatches 152; Indels 68; Gaps 13;

Qy 347 KISLDFQDVEIRTIQILAKESGMNIVASDSVNGKMTLSLQVPMQDQALDLMQARN-LD 405  
 Db 926 LUTINAEDAPLYDLLEETASELIGISVMFVSPSEKITWKADNVAWEKFIDILISQYGLP 985  
 Qy 406 MRQQNINVIAPRDELLAKDAFLQAEKDIALGALYSQNLQYKYNVEFRSILRLDNA 465  
 Db 986 DNKNGVYVVKPKQDL-----ARRIYDV-----PHNFD-QIKALIEFVG----- 1024  
 Qy 466 DTGHNLTIVSGRGLVDPATNTLIVTDRSVIEK-FRKLIDELDVPQVQMTIARIVE 524  
 Db 1025 -----GTVYVSLNNFVVTGISETIKRELQNLIERKUKPKQIEISAKIVD 1071  
 Qy 525 AA--DQGRDLGVKFGATGKKLKNDSAFGNGVNSGFGGDDKMGAETKINLPITAAANS 582  
 Db 1072 RSLIDLSKETGLD-LGE-----NVNVGSSG-----AESFSVTDYLDLF 1110  
 Qy 583 ISLVRAISSGALNLSASESLSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANGG 642  
 Db 1111 EXIFGEILNLTSLQFSQDKTNTLDDILASPRIVTTSKGEARILIGDRIPYVTDPT--NGD 1168  
 Qy 643 SSTNTELKAVLGLTVTPNITPDGQIIMTVKINKDSPAQCASGN--QTILCISTKNLNTQ 700  
 Db 1169 GTPEVQFLETGIELSITPFVRSDDTIELDLFYKASEP-----GNVINEVPGERTEAQTH 1223  
 Qy 701 AMVNGOTLIVGGIYEDNGNTLTQVPLGIDPVGNIPLFKRGKKTDRRELLIFITPRIM 760  
 Db 1224 LIVKNGSTITIGLIREVNTVTSKLPFLGLDLPVIGQFFRTKSENKPKRDLVIFLTVRW 1283

A:Title: Molecular cloning and characterization of 13 out genes from Erwinia carotovora bacteria.  
 A:Reference number: S32857; PMID:93316842; PMID:8326859  
 A:Accession: S32858  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-649 <REE>  
 A:Cross-references: EMBL:X70049; NID:G42184; PIDN:CAA49645.1; PID:G42186  
 C:Genetics:  
 A:Gene: outD

Query Match 9.0%; Score 346.5; DB 2; Length 649;  
 Best Local Similarity 23.4%; Pred. No. 2.6e-12;  
 Matches 117; Conservative 84; Mismatches 189; Indels 109; Gaps 13;

Qy 284 TPQVKYTLKRLNNDTOLIITTAGNWELV-----NKSAAPGYFTFQVLPKPKONLES 333  
 Db 171 TIVERVQGTGDRNVTTPISYASSTEVKWNELNKMDEKSPALPGMLTANVADERTNSA 230  
 Qy 334 GGVNAPKFTTGRKISLDQDVEIRTIQILAKESGMNIVASDSVNGKMTLSLQVPMQ 393  
 Db 231 AG-----FGEENSR-----OR 241  
 Qy 394 ALDVMQARNLDMRQ--QGNIVNIAPRDELLAKDAFLQAEKDIALGALYSQNFOLKYK 451  
 Db 242 VIDMVKQ---LDRQQAVQGN-----TKVIYLKVA 267  
 Qy 452 NVBEFRSILR-LDNADTTGNRNTLVSGRGSVLII--DPATNTLIVTDRSVIEKPKLIDE 508  
 Db 268 KAADLVEVLTVGDSIQTDQNALPALRKDISIKAEHQTNLSLVNAAPDIMRQLEGVIAQ 327  
 Qy 509 LDVPAQVQMTIARIVEADGFSRDLGVKFG--ATGKKLKNDSAFGNGVNSGFGGDDKW 566  
 Db 328 LDIRRPOVLVEAIIAEVQDAGNVLGVQWANKNAGVTQTNT---GLPITTMAGADQF 383  
 Qy 567 GATKINLPITAAANSISLVRA-ISSGALNLSASESLSKTKTLANPRVLTQNRKEAKI 625  
 Db 384 RRDGTLGTAATTALGGFNGIAAGFYQGNMGLMTALSSNKNKDILATPSITVTLDNMEATF 443  
 Qy 626 ESGYEIPFTVTSIANGSSNTTELKAVLG--LTVTPNITPDQIIMTVKINKDSPAQCA 683  
 Db 444 NVGQEVFLVLAGSQTSGDNVQFVERKTGKIKVQPNEGSDVLEIEQEVSSVADAA 503  
 Qy 684 SGNQTLI--CISTKNLNTQAMVNGOTLIVGGIYEDNGNTLTQVPLGIDPVGNIPLFK 741  
 Db 504 SSSSTNLGATFNRTVYNNAVLVSSGDTVVVGGLLDKSTNESANKVPLGLDIPVLGYLFRS 563  
 Qy 742 RGKKTDRRELLIFITPRIM 760  
 Db 564 NSTETKRNLMFLIRPSII 582

RESULT 19  
 S39653  
 xcpQ protein - Pseudomonas aeruginosa  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 31-Dec-2000  
 C:Accession: S39653; A83259  
 R:Akrim, M.; Bally, M.; Ball, G.; Tommassen, J.; Teerink, H.; Filloux, A.; Lazdunski, M. Microbiol. 10, 431-443, 1993  
 A:Title: Xcp-mediated protein secretion in Pseudomonas aeruginosa: identification of t  
 A:Reference number: S39652; PMID:95020542; PMID:7934833  
 A:Accession: S39653  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-658 <AKR>  
 A:Cross-references: EMBL:X68594; NID:G431183; PIDN:CAA48582.1; PID:G431185  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Li ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pat  
 A:Reference number: A82950; PMID:20437337; PMID:10984043

A;Cross-references: GB:AE000697; NID:92983212; PIDN:AAC06820.1; PID:92983222; GB:AE000  
A;Experimental source: strain VFS  
C;Genetics:  
A;Gene: aq\_585

Query Match 8.7%; Score 336; DB 2; Length 705;  
Best Local Similarity 22.4%; Pred. No. 1.2e-11;  
Matches 128; Conservative 110; Mismatches 206; Indels 128; Gaps 21;

QY 255 QCHDHIIVTLKNTLPTTLQSL-----DVADFKTPVQKVTLLKLNND 297  
DB 193 REVDLIIVKLSNNLIKISKETLAFDVEGDVSSINKLSIKQVTSAPKLYDK-DLG 251  
QY 298 TQIIITAGNWE-----LVNKSAAQGYFTFQVLPKKQNLSEGGVNNAPKFTTG 345  
DB 252 KIMVIDMAENIEKRLDRVLDLIELMSRETTG-----EKEKSET-----TPR----- 295  
QY 346 RKISLDFQDVEIRTIQILAKESGMNIVAS---DSVNGKMTSLKDVDPWDQALDLMQAR 402  
DB 296 -----EIEKVFYFKNKRDLIALSRLENFSGEVILNI-----DKDFNALIVTS 340  
QY 403 NLD-MRQOGNIVNIAPDELLAKDKAPL-----OAEKDIAIDLALYSQ 444  
DB 341 NRSVIKSVGTLL-----KDLTESIDKAYLITKIPYRISPYELKKIEPMLSEVGYT- 395  
QY 445 NFOLKYKNVEEPRSIIRLDNAD--TTGNRNTLVSGRSLVIDPATNTLIVTD-----TRSV 498  
DB 396 ---LSVNTDEKELISYKNTPPATAPNECTLEKRAFFV--PENNAILIKDYPERIEKI 450  
QY 499 IEKPRKILDELDPQAQVMIEARIVEAADGFSDLGVKFGATGCKKLKNDTSAFGMGWS 558  
DB 451 REKFKFLSEKPI---KIKIRAKLVEVEKSLRELGLISWRTVFSKAVI----- 495  
QY 559 GFGDDDKWGAETKINLPITAAANSISLVAISSGALNL---ELSAESLSKTKTLANPRV 615  
DB 496 ---PFWQGETAPRTVTPGQPSGLITFTFORNLNLEPKLLAYBQEGRAKNVAESYV 551  
QY 616 LTQNRKAKIESGVEIIFTVTSTANGSSNTTELKAVLGITVTPNITPDGQIMTV--- 672  
DB 552 ITVNGEPAVSSGIEFFPTEVSLSGGIANPEKYESIPVLTTPVLPDGNILSLTYLA 611  
QY 673 --KIN--KDSAPQACASGNQITLCISTKNLTQAMVENGTLIVGGIYEEDNGNTLTQVPL 728  
DB 612 RQINSVQEFV--TQTLTKIPVESTRIDVKIPKNGETWIGGAVEKSDSITESGVPK 670  
QY 729 LGDIPVIGNLFTKTKTDRRELLIFITPRIM 760  
DB 671 LREVPILGLMFKTQTKQLRDRELLIFITPEII 702

## RESULT 21

general secretory pathway protein D precursor XF1527 [imported] - Xylella fastidiosa  
B82671  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: B82671  
R:Anonymous; The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq  
Nature 406, 151-157, 2000

A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A;Reference number: A82515; MUID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: B82671

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-775 <SIM>

A;Cross-references: GB:AE0003982; GB:AE0003949; NID:g9106554; PIDN:AAF94336.1; GSPDB:GN  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carr  
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La

A;Accession: A83259

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-658 <STO>

A;Cross-references: GB:AE004734; GB:AE004091; NID:g9949204; PIDN:AAC06493.1; GSPDB:GN001

A;Experimental source: strain PA01

C;Genetics:  
A;Gene: xcpQ; PA3105

Query Match 8.8%; Score 337; DB 2; Length 658;  
Best Local Similarity 23.6%; Pred. No. 9.5e-12; Indels 148; Gaps 28;  
Matches 147; Conservative 117; Mismatches 210;

QY 235 KNAGIIEAALGCFAGQPDISOQHDIIVTLKNTLPTTLQSLDVAADFKTPVQKVTLLKRL 294  
DB 35 ENSG-----GNAPVAGNQOEAWHTNLKD-----ADIREFIDQISEITGETF 77  
QY 295 NNDTQIIITAGNWEVNYK---SAPGFTF-----QVLPKKQ-NLE 332  
DB 78 VVDPV---KQGVSVSKAQLSLSEVQLFSLVNSTGFTVVAOGDQARIVPAEAKTE 133  
QY 333 SGVNNAPKFTGRKISLDFQV--EIRTIQILAKESGMNIVASDVNGKMTSLKDVPM 391  
DB 134 AGGQSAPRLETRVIOQSPVSELPLIRPLVPQYG-HLAAVFSANA-LIISDRSANI 191  
QY 392 QDALDLMQARNLDRQOG---NIVN-----IAPDELL-----AKDKAFLOAKD 434  
DB 192 ARIEDVI---RQID--QKSGHDSVINLRYGWMDAAEVNMMRSGQAKGAAGQAVIAD 246  
QY 435 IADLALYSQNLQKYNVEEPRS-----ILRLDNAD-----TTGNR 471  
DB 247 ARTNRLLILGPPQARAKLVLAQSLDTPARSANTRVIRLRENDAKTLAETLQIGSEGK 306  
QY 472 NTLVSGRG-----SVLI--DPATNTLIVTDSVIEKFKLIDELDPQAQVMIE 519  
DB 307 NN--GGQGEQTGGGRPSNILIRADESTNALVLLADPTVNALEDIVRQLDVPRQAQVIE 364  
QY 520 ARTVEAADGFSRDLGVKFGATGCKKLKNDTSAPGMGN--SGFGDDDKWGAETKINL--- 574  
DB 365 AAIVEISGDIQDQAVGQ-----WAINKGSGGKTNTFANTGLSIGTL 407  
QY 575 -----PITAAANSISLVAISSGALNLSELSKTKTLANPRVLTQNRKAKIE 626  
DB 408 QSLNESKAPESIPDGAIVGIGSSSFGALVTALS---TKSNLLSTPSSLTLDNQAKEL 464  
QY 627 SGVEIFP---TVTSIANGSS--TNTLKKAVLGITVTPNITPDGQIMTVKINDSPA- 680  
DB 465 VQQRVPFTQSGYTNSESSNPFTTVERKDIGVSLKVTPHN-DG-AAALRLIEQEISAL 522  
QY 681 --CCASGNQITLCISTKNLTQAMVENGTLIVGGIYEEDNGNTLTQVPLGIDIPVIGNL 738  
DB 523 LPNAQQRNNTDLITSKESIKSTILAENGQVIVIGGLIQDDVSOAESKVPPLGDIPLGLRL 582  
QY 739 FKTRGKKTDRRELLIFITPRIM 760  
DB 583 FRSTKQTHTKRNLMLVFLRPTIV 604

## RESULT 20

conserved hypothetical protein aq\_585 - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999

C:Accession: F70352

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O

V.

A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A;Reference number: A70300; MUID:98196666; PMID:9537320

A;Accession: F70352

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-705 <AQF>

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B.  
A;Authors: Martins, E.M.F.; Matsumura, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A;Reference number: A59328  
A;Contents: annotation  
C;Genetics:  
A;Gene: XF1527

Query Match 8.7%; Score 333.5; DB 2; Length 775;  
Best Local Similarity 20.4%; Pred. No. 1.9e-11;  
Matches 176; Conservative 126; Mismatches 275; Indels 287; Gaps 31;  
QY 28 TDIKVSSLPNKQKIVKVSF-----DKEIVNPTGFTVSSPA- 62  
DB 39 TDIPTTQLPDVAGTATVLPDPTTETPTALLSDAGCPLEVRGNGKVINQTVAAATPPSM 98  
QY 63 ---RIADFEQTGISMQQVLEADPLSK-----ISAAQNSRA 99  
DB 99 GVAGKSGATFNFEGLSQAANKAILGDMQNYFIASGVQGTVLSTPKPVSSAQAALL 158  
QY 100 RLVLNLPKQGVNTEVRGNKWIIFNESDDTVSAPARPAVKAAPAKAQGGQRTVTVQR 159  
DB 159 EMVLGN-----NARMINNGRYSIVQDQALAGTVAEST-----APPAVARG-----FEVR 205  
QY 160 SIRIQLYPGKTTAAAPFTESVSVSAPESPAKQQAASAKQQTAAAPAKQTAAPAKQQA 219  
DB 206 VV-----PLKYSISASEMK-----VLDPY 224  
QY 220 AAPAKQTNIDFRKQGNAGIIEAALGAGAGQDFISQQHDIIVTLKNTLPTTLQRLSDV 279  
DB 225 ARPNAIVSIDTRN-----LITLAGTR-----VELENYL-----RTVQI 258  
QY 280 ADFK-----TPQVKTLKRLANDQLIITTAGNWLKGAAPGVFTFQVLPKKQN 330  
DB 259 FPDVWLSGMSGVFPQSGKADKVAADLEKVFGESE-----KTPSAGMFRP--MP----- 306  
QY 331 LESGNNVNAKPTFTTKISLDPQDVEIRTLQILAKESGMNIVASDSYNGKMTL---SLK 387  
DB 307 LESA---NAVLTTPQASYLD---QIKWLD-----SVDSVGGALSLSFVALK 348  
QY 388 DYPWQALDLYNQARNLDMRQGNLVNTPARDELLAKDKAFLOAEKDIADIGALYSQNFQ 447  
DB 349 YIKANDLANRLTEVCGARSDNSVSLAPG-----AQGLVILGSGSG 391  
QY 448 LKYKNVEFRSILRLDNADTTGNRLTVSGRG--SVLIDPAT----- 487  
DB 392 DSLPSAGGGSLSAVPSNGGTDNTSSANGGLCGSTLQSPRTQNGSVTLHWQGDTVGVS 451  
QY 488 -----NTLIVTDRSVIEKFKLIDELVPAQOVMIEARIVEAADGFSRDLGVKFGATGK 542  
DB 452 AVEEINTULVRATPOAWRSIRDVIEKLDVMPNQVHIEAQVAEVS----- 495  
QY 543 KKLKNDTSAFGWGNSGFGGDDKWAETKINLPITAAANSISLVRAISSGA----- 593  
DB 496 ---LTNQLS---YGVN-----WFFQNSVNAADAADNGASNGTGIGLAGLPSAAGRS 542  
QY 594 -----LNLELSASSELSTKTKLANPRVLTQNRKEAKIESGVEI 631  
DB 543 GWKSIAGKVTNSGLAWTFGLGNAAAIINALQVOTQVRLLOTPSVFVRNNAEATLNVGARI 602  
QY 632 PFTVTSIAN--CGSSTNTLKKAVLG--LTVTPNITPDG---QIIMTVKINKDSPAOCA 683  
DB 603 PINSTISINTGLGNSYSSVQIDTVGVILKVRPRVTKQGMVFLDIVQEVSTFGSLPAACS 662  
QY 684 SQNTIL---C---ISTKNLNTQAMWNGGTLVGGIYEEDNGNTLTKVPLLDGIPVIG 736  
DB 663 SASSTLVNSAACNVQINTERIKTEAAVSGGDTIMLAGLIDNSNGKSGNSGVFPLSKVPIVG 722  
QY 737 NLFKTRGKTKDRRELLIFITPRIM 760

Db 723 GLFGTKAQNRRREIIVLTPEIV 746  
RESULT 22  
H65125  
probable general secretion pathway protein d precursor - Escherichia coli (strain K-12)  
C;Species: Escherichia coli  
C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C;Accession: H65125  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
A.; Rose, D.J.; Mau, B.; Shaoh, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of Escherichia coli K-12.  
A;Reference number: A64720; MUID:97426617; PMID:9278503  
A;Accession: H65125  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-654 <BLAT>  
A;Cross-references: GB:AE000409; GB:U00096; NID:G1789718; PIDN:AAC76350.1; PID:G178972  
A;Experimental source: strain K-12, substrain MG1655  
C;Genetics:  
A;Gene: ynfE

Query Match 8.6%; Score 332; DB 2; Length 654;  
Best Local Similarity 21.5%; Pred. No. 1.8e-11;  
Matches 123; Conservative 97; Mismatches 183; Indels 168; Gaps 17;  
QY 351 DFODVEIRTLQILAKESGMNIVASDSYNGKMT-----LSLKD----- 388  
DB 35 NFNADIRQFVEIGHGHLGKTLIDFSVQGTISVRNDTFSQEYVQFFLSILDLGYGSV 94  
QY 389 -----VPMDQ--ALDLVMOARN-LDM 406  
DB 95 ITLDNGFLKVRGANVKTSPGMIADSSRPGVGBDELVTRIVPLENVDPARDLAPLRQMDA 154  
QY 407 RQGNIVNTAPRELLAKOKA-----FLOAEKDIADIGALYSQNFOLKYNKVEFSILR 461  
DB 155 GSYGNVHVHPEPSNVLTGTGRASINKLIEVKRVGVIGTEKQOIIEHLEVASAEDLAELN 214  
QY 462 -----LDNADTTGNRNT---LVSG----- 477  
DB 215 QLISESHGKXQMPALLSAKIVADKRTNLSLIISGPEKARQITSLLSKSLDVEESEGNTRV 274  
QY 478 -----RG-----SVLIDPATNTLIVTDRSV 498  
DB 275 YILKYAKATNLVBLTGVSEKLDERGNARFKPSSGAMDNVAITADEQTNLSLITADQSV 334  
QY 499 IEKFRKLIDELVPAQOVMIEARIVEAADGFSRDLGVKFG--ATGKKLKNDTSAFGWGV 556  
DB 335 QEXLATVIAERLDIRRAQVLVEAIIIVEQDGNGLNLGVQWANKVGAQQFTN--TGLPIFNA 393  
QY 557 NSGFGDDDKWAETKINLPITAAANSISLVRAISSGALN---LELSASSELSTKTKLAN 612  
DB 394 AQGVADYKXNGGITSAN-----PAMDVFSAYNGVAAGFNGDWGVLLTALASNNKNDILAT 449  
QY 613 PRVLTQNRKEAKIESGVEIPFTVTSIANGGSSNTTELKKAVLG--LTVTPNITPDGQIIM 670  
DB 450 PSIVTLNKLASFNVQDVPVLGSGSTTSQDNTVNTVERKTGTVKLTPTQVN--EGDAVL 508  
QY 671 TVKINKDSPAQASGNOTI--LCISTKNLNTQAMWNGGTLVGGIYEEDNGNTLTKVPL 729  
DB 509 -LEIEQVSSVDSSNSSTLGTPTNTRTIQNAVILVKTGETVVLGGLLDDFDSKEQSVK 567  
QY 730 GDIPVIGNLFKTRGKTKDRRELLIFITPRIM 760  
DB 568 GDIPVQQLFRYTSSTERAKNLMVFIPTII 598

RESULT 23  
B47021  
pectic enzyme secretion protein OutD - Erwinia chrysanthemi  
C;Species: Erwinia chrysanthemi  
C;Date: 24-Feb-1994 #sequence\_revision 18-Nov-1994 #text\_change 18-Nov-1994

C:Accession: B47021  
R:Lindeberg, M.; Collmer, A.  
J. Bacteriol. 174, 7395-7397, 1992  
A:Title: Analysis of eight out genes in a cluster required for pectic enzyme secretion b  
A:Reference number: A47021; MUID:93054355; PMID:1429461  
A:Accession: B47021  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-712 <LIN>  
A>Note: sequence extracted from NCBI backbone (NCBIP:118271)

Query Match 8.6%; Score 330; DB 2; Length 712;  
Best Local Similarity 27.9%; Pred. No. 2.7e-11;  
Matches 86; Conservative 80; Mismatches 126; Indels 16; Gaps 8;

QY 464 NADTTGNRTLVSGRG-SVLIDPATNTLIVTTRSVIEKFKLIDELDVPAQQVMIEARI 522  
DB 347 SANSSGRRIVTQGEKVTVRANDQNSLIITRPDMDRLDQVINGLDIRRQVLEAII 406  
QY 523 VEAADGFSRDLGVKFGATGKKLNDTSAGFGVNSGFGDDKGAETKINLPITAAANS 582  
DB 407 AEIQDADGLNLGIQW--ANKRAGMTQFTNTGIPISAMIGTQDFRSDGTLT---TAYASA 461  
QY 583 ISLVRAISSGAL---NLELSASELSKTKTLANPRVLTONRKEAKIESGYEIPF---TV 635  
DB 462 LSNFNGITAGFYRGNWSMLLTALSSDGKNDVLATPSIVTLDNMEATFNVGQVPLTGSQ 521  
QY 636 TSANGSSNTTELKAV-LGLTVTNITPDGQIMTVKINKDSPAQ-CASGNQTI-LCI 692  
DB 522 TTVSGDNIENFVTKVTGFKLRVKPQINEGDSVLLQIOEVSVAEGNSNSLGVTF 581  
QY 693 STKNLNTQAMVNGGTLIVGGIYEEDNGNTLTIVKPLLDGIPVGNLTKRGKKTDRRELL 752  
DB 582 NRTVNNAVMTNRETIVVGGLLDKTAETNNKVPVLLGDIPLWLSIFRSKTOFTMSKRLM 641  
QY 753 IFITPRIM 760  
DB 642 LFLRPTII 649

RESULT 24  
S64727  
Protein secretion protein xcpQ precursor - Pseudomonas putida  
C:Species: Pseudomonas putida  
C:Date: 06-Dec-1996 #sequence\_revision 06-Dec-1996 #text\_change 21-Jul-2000  
C:Accession: S64727; S47504  
R:de Groot, A.; Kijger, J.J.; Filloux, A.; Tommassen, J.  
Mol. Gen. Genet. 250, 491-504, 1996  
A:Title: Characterization of type II protein secretion (xcp) genes in the plant growth-  
A:Reference number: S64727; MUID:96185881; PMID:8602167  
A:Accession: S64727  
A:Molecule type: DNA  
A:Residues: 1-591 <DEG>  
A:Cross-references: EMBL:X81085; NID:G3293032; PIDN:CAA56979.1; PID:G531740  
C:Genetics:  
A:Gene: xcpQ  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-591/Product: protein secretion protein xcpQ #status predicted <MAT>

Query Match 8.6%; Score 329.5; DB 2; Length 591;  
Best Local Similarity 23.6%; Pred. No. 2.2e-11;  
Matches 123; Conservative 92; Mismatches 174; Indels 133; Gaps 19;

QY 350 LDFQDVEITLQILAKESGNIVASDSVNGKMTLSKDVDPWDQALDLVMOARNLNRQ 409  
DB 39 VNFVDTELGEFDSVSRITGTTFFVDPRVKGVIVTVLDLHDAIDVIDFLAQ---LRAQ 95  
QY 410 G-----NIYNIAP---RDELLAKDKAFLOAE-----K 433  
DB 96 GYATVDLPNGSVKIVPDQAAARLEPVPEPAGQOEGSDSVATRVFNVAATEOVLGILK 155  
QY 434 DTAD--LGNL--YSQNFOLKY----KNVEEFSILR-LDN-ADTGNRLTV----- 475

Db 156 PLIDPRGVITPYPAHQVLVWTDWRSNLERIASLLRQLDRPSETAGSSSTQVIYLRHATA 215  
QY 476 -----SGRGSVL-IDPATNTLIVTDRSVI 499  
Db 216 SEYVKVLRGLSQEAPAEAGVAGEAKDRPWMAAPGSGIRLIEYEGTNAVVMVGPDSSEL 275  
QY 500 EKFRKLIDELDVPAQQVMIEARIIVEAADGFSRDLGVKFGATGKKLNDTSAGFGW-VNS 558  
Db 276 AAYRAIVEQLDIRRAQVVVEAIIAEVSDSAQELGVQWLFADEK-----FCAGIVNF 327  
QY 559 GFGDDKGAETKINLPITAAA-----NSISLVRAISS-----GALNLEL--SA 600  
Db 328 GSG-----VNIAIAGAAAGDETEALGALLSAGATAGIGHFGGFGFPAFLVNA 378  
QY 601 SESLSKTKTLANPRVLTONRKEAKIESGYEIPFTVTSIA--NGSSSTNTTELKAVLGLTV 658  
Db 379 LKXSGGNLLSTFTLLLDNAEASILLVQEVFPFVTGSVTQNNANPYQTIERKEVGVKLRI 438  
QY 659 TPNITPDGQIMTVKINKDSPAQCSNQITLICTKNTQAMVNGGTLIVGGIYEED 718  
Db 439 KPQINIDNSVRLDIVQEVSSIAADSDVI--TNKREIKTKVMVEDNGLIVLGLLSIDE 496  
QY 719 NGNTLTVPVLLGDIPLVGNLTKRGKKTDRRELLIFITPRIM 760  
Db 497 LSTSNQRPVLLGDIPLVGNLTKRGKKTDRRELLIFITPRIM 760

RESULT 25  
S28014  
OUTD protein - Erwinia chrysanthemi  
C:Species: Erwinia chrysanthemi  
C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 21-Jul-2000  
C:Accession: S28014; S23886  
R:Condemine, G.; Dorel, C.; Hugouvieux-Cotte-Pattat, N.; Robert-Baudouy, J.  
Mol. Microbiol. 6, 3199-3211, 1992  
A:Title: Some of the out genes involved in the secretion of pectate lyases in Erwinia  
A:Reference number: S28011; MUID:93086427; PMID:1453958  
A:Accession: S28014  
A:Molecule type: DNA  
A:Residues: 1-710 <CON>  
A:Cross-references: EMBL:X65265; NID:G3152953; PIDN:CAA46370.1; PID:G581156  
C:Genetics:  
A:Gene: outD  
A:Start codon: GTG

Query Match 8.5%; Score 328; DB 2; Length 710;  
Best Local Similarity 27.4%; Pred. No. 3.5e-11;  
Matches 85; Conservative 77; Mismatches 134; Indels 14; Gaps 7;

QY 464 NADTTGNRTLVSGRG-SVLIDPATNTLIVTTRSVIEKFKLIDELDVPAQQVMIEARI 522  
DB 347 STSSSGRTITIQGEKVTVRANDQNSLIITAPPDMDRLDQVINGLDIRRQVLEAII 406  
QY 523 VEAADGFSRDLGVKFGATGKKLNDTSAGFGVNSGFGDDKGAETKINLPITAAANS 582  
Db 407 AEIQDADGLNLGIQW--ANKRAGMTQFTNTGIPISAMIGTQDFRSDGTLT---TAYASA 461  
QY 583 ISLVRAISSGAL---NLELSASELSKTKTLANPRVLTONRKEAKIESGYEIP-FTVTS 637  
Db 462 LSNFNGITAGFYRGNWSMLLTALSSDGKNDVLATPSIVTLDNMEATFNVGQVPLTGSQ 521  
QY 638 IANGSSSTTELKAV-LGLTVTNITPDGQIMTVKINKDSPAQCSNQITLICI--ST 694  
Db 522 TTSADNIFNTVERKTVGILRVKPKQINEGDSVLLQIOEVSVAEGNSNSLGVTFNT 581  
QY 695 KNLNTQAMVNGGTLIVGGIYEEDNGNTLTIVKPLLDGIPVGNLTKRGKKTDRRELLIF 754  
Db 582 RTVNNAVMTNRETIVVGGLLDKTSVESNDKVPVLLGDIPLWLSIFRSKQEVKRNLMF 641  
QY 755 ITPRINGTAG 764  
Db 642 LRPTIIRDPG 651

| Query Match           | 8.4%             | Score 322.5   | DB 2       | Length 642 |
|-----------------------|------------------|---|------------|------------|
| Best Local Similarity | 22.3%            | Pred. No. 6.2e-11   |            |            |
| Matches 128           | Conservative 104 | Mismatches 218  | Indels 125 | Gaps 21    |
| Qy                    | 242              | LAALGFAGAGPDIISQOHDHIIIVTL-----KNRTLTPTTLQRLS-----DVADEKTPVKQVT-- | 290        |            |
| Db                    | 72               | LDVYGA-----VDMHNGILKVRSKDAKTSAPVVASDVSPGTGDEVVTRVPVSNVAAR         | 127        |            |
| Qy                    | 291              | -----LKRNNDT--QLIITTAGNWEIKNKSAAAGVFTFQVLPKKONLBSGGVNNAPKT        | 342        |            |
| Db                    | 128              | DLAPLLPQLNDNAGAGSVHVEPSVLLMTGTAAVMKRLMEIVERVVDKGNRSVATVPLT        | 187        |            |
| Qy                    | 343              | FTGKTSISLDFQDVEIRILOILAKESGMNIV-----ASDSVNGKWTLSLXDVPWDQA         | 394        |            |
| Db                    | 188              | YASA-----TDVABRLVTELTAKTAPAWMTAKLVADERTNSVLVSGSPIISOORI           | 239        |            |
| Qy                    | 395              | LDLVMQARNLDMRO--QGNIVNIAPRDELLAKOKAFLQAEEKDIADLGALYSQNFQKLYKN     | 452        |            |
| Db                    | 240              | ISIIKQ---LDRQEDVQGN-----TKVILYLYKAK                               | 265        |            |
| Qy                    | 453              | VEBERSIL-----RLDNADTGTNRTNLVUSRG-SVLIDPATNTLIIVTDRSVTEKFERKLI     | 506        |            |
| Db                    | 266              | AKDLVEVLITGISSSIEH-DSKKSPSTALRKGVITKSGEQTNALITGAPDVIIRDLENVI      | 324        |            |
| Qy                    | 507              | DELDVPAQAWYIEARIVEAADGFSRDLGVFGATGKKKLKNDTSAFQGVNSQFGGDDKW        | 566        |            |
| Db                    | 325              | SQDIRRPQVLVEAIIAEIQADGLNIGIQ-----W-VNKIAG-----V                   | 362        |            |
| Qy                    | 567              | GAETKINLPITA-----AANSISLVRAISSGALN-----LELSASESLSKTKT             | 609        |            |
| Db                    | 363              | AQFTSTGLPITTWQTRQNEILDSQSNALSMFNGIAAGFYQGVNWMALLTALSTSSKNDI       | 422        |            |
| Qy                    | 610              | LANPRVLTONRKEAKIESGVEIP-FTVTSIANGSSNTTELKXAV-LGLTVTNITPDGQ        | 667        |            |
| Db                    | 423              | LATPSIVTLDNMEATFNVGQVEPVLGSGQTTSGDNIENFTVERKTVGIKLRVKPQINEGDS     | 482        |            |
| Qy                    | 668              | IIMTVKINKDSPQACASGNOTIL--CJSTKNLNTQAMVNGGTLIVCGIIVEEDNGNTLTIX     | 725        |            |
| Db                    | 483              | VLLIEQVEGSGVADTAVATTTDLGATNTRTVTVNMLVNGETVVVGGGLLCKSIRGSEK        | 542        |            |
| Qy                    | 726              | VPLGDIPIVIGNLFPKTRGKXTDRRELIIFITPRIM                              | 760        |            |

| Query Match           | 8.3%;             | Score 319.5;   | DB 2;       | Length 687; |
|-----------------------|-------------------|--|-------------|-------------|
| Best Local Similarity | 21.5%;            | Pred. No. 1e-10;   |             |             |
| Matches 148;          | Conservative 115; | Mismatches 259;  | Indels 167; | Gaps 24;    |
| QY                    | 183               | SVSAPTSAPAKQAAASAKQOOTAAPAKQOOTAAPAKQOAA-----AAPAKQTNIDFRKDGKN | 236         |             |
| DB                    | 9                 | SVLAAVAALVMGASL---GPAPVLAQTCLLVNQDADIRVFQDVAKSTGTTFIIDPRV      | 65          |             |
| QY                    | 237               | AGHIEALAGFAGQPDISOQHDHIIIVTLKNHTLPTTLQRLSDLVADPKVPQKVTLLKRLNN  | 296         |             |
| DB                    | 66                | KGTVTVASNGPLNRREL---FEVFLATL-----RANN                          | 94          |             |
| QY                    | 297               | DTQLIITTAGNW-----ELVNKSAAPGYFTFOVLPKKONLESGGVNNAPKTFTGRK       | 347         |             |
| DB                    | 95                | ---FVATPAGNGAYRIEPPSENAARQSPSAGGQFATVP-RLRLTDAASAVEMKLPLVGPQ   | 150         |             |
| QY                    | 348               | ISLDFQDVEIRTIILQILAKESGNMIVASVNSGKMTLS--LKDVPMDQALDLVMQARNLD   | 405         |             |
| DB                    | 151               | -----GVVANPRGAVVVVADYADNVRIRGLLAQVQDQDRAVMHTVTTLTHSS           | 197         |             |
| QY                    | 406               | MRQCGMIVN-----IAPRELLAKAKAFQAQKD-----IADLG                     | 439         |             |
| DB                    | 198               | ARIEAQLNDMLATPGAEGSKSGRGAVTVVPD---SSNSVLLRGDSGAVORLLPFIADLD    | 254         |             |
| QY                    | 440               | ---ALYSQNFQ---LKYKNVEEFSILR-----LDNADTTGNRNTLVSG               | 477         |             |
| DB                    | 255               | PRAESSDDVRVPLRHANAEMQLPVLQVLGQATTSVTTSSSRGLSNARTATCAGTATTA     | 314         |             |
| QY                    | 478               | -----RGSVLIDPATNTLITVDTFSVIEKPKRLIDELDPVPAQQWKEARIV            | 523         |             |
| DB                    | 315               | PSAPVAATSPANGQPQATIRAFPGCANALINAPPETORTLAEVIRQLDVRREQVLVEAIVV  | 374         |             |
| QY                    | 524               | EADGFSRDLGVKF--GATGKKKLKNDTSFAGWGVN-----SGFGGDDKMGAAET         | 570         |             |
| DB                    | 375               | EVTDGTAQQLGVQFLLGCT-----NGTIPF-VATVNTNATPSLLPLVGAATTTSGADT     | 427         |             |
| QY                    | 571               | KINLPI-TAAANSISLVRAISSGALNE-----LSASESLSKYTLANPVLVLTQNRK       | 621         |             |
| DB                    | 428               | EALKSRDAAVSLLTFASGVTSGVGRSGDALFGAIINAVKDTGKNLLSTPSIMTLONE      | 487         |             |
| QY                    | 622               | EAKIESGYRIPFTVTSIANGSGSTN---TELKKAVALGLTVTNITPDGOIIMTVKINKD    | 677         |             |
| DB                    | 488               | EARILVQGEVPIITTEVL-GDSNANPRTIQRQNVGILEKVPQINAGGGITLPLRQEV      | 546         |             |
| QY                    | 678               | SPAQASAGNQTLICISTKNLNTQAMVNGGTLIVGGYEBDNGNTLTKYPLLLGDIPIVGN    | 737         |             |
| DB                    | 547               | SVAGPVSVGSESLINKREIETTALVDGDDIVVLGGLLDQOETQSQARPTPLGDPVCLGA    | 606         |             |
| QY                    | 738               | LFKTRGKKTDRRELLIFITPRIMGTAGNS                                  | 766         |             |



Db 607 LFRSTARERKTNLMVFIRPRISSADA 635  
RESULT 28  
E70411  
General secretion pathway protein D - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
R:Accession: E70411  
R:Beckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Oviatt, D.; et al.  
V: Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: E70411  
A:Molecule type: DNA  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Cross-references: 1-625 <AOQ>  
A:Cross-references: GB:AE000732; MID:G2983704; PID:AA07271.1; PID:G2983708; GB:AE000655  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: gspD  
Query Match 8.3%; Score 318.5; DB 2; Length 625;  
Best Local Similarity 23.3%; Pred. No. 1e-10;  
Matches 143; Conservative 95; Mismatches 246; Indels 129; Gaps 17;  
QY 215 AKQAAAPAKATIDFRKDGK-----NAGIIEAALGAFAGQPDISQOHDHIIIVTLKNHTL 269  
Db 31 AVGVAKLLGKLVDPVKVKITIIISNGEISESEALELFSQALASQGSLLI----- 81  
QY 270 PTLQRLSDVADFTPVQKVTLKNNDTQLIITTAGNMELVNKSAAPGYFTFQVLPKKQ 329  
Db 82 -----LEKDTWKIIPASOGYPFTEIKAGKGGEFTLVLYKUK 117  
QY 330 NLESGGVNNAKPT-----TGRKISLDFQDV--EIRTIQLLAKESGMNIVAS 375  
Db 118 NTNASQVVSALRPPLSPYGRIFAHQAQNSVITDYADSVNKKILFYL----- 166  
QY 376 DSVNGKXTL-SLKDPVWDQALDLMQARNLDMRQOQGNVNIAPRDELLA----- 423  
Db 167 DSGGGEVRVYKLYKVPYVYVKKLNPFTSLTKRYGEPVVIAGVDENAVAVYANKEIHR 226  
QY 424 -----KQKAFLOAEKDIALGALYSQNFQKLYKNVEEFRSILRLDNADTTGNRN-- 472  
Db 227 IIEKVISDLDDPSSLEAERSF-----YIIPNFVSABEIIYESLNSVFKGKSVGTTKAK 282  
QY 473 -----TLVSGRGSVLIDPATNLTIVTDRSVIEKFKLIDELDPVPAQQWIEAR 521  
Db 283 KKGQVQALTSITLKSQM-KIGFDKXTNLSLILYATKSEYEAVERFIKKIDKRRKQLLTAT 341  
QY 522 IVEAADGFSRLGVKFGATGKKLKNDSAPFGWNGSGFGGDDKKGAEETKINLPITAAAN 581  
Db 342 IIEASAKSILEAGIRWQILG-----THGGAAPKGSLLQDVYNAIKSGNFVIGGFSK 392  
QY 592 SISLVRALSSGALN-----LELSASESISKTKTLANPRVLTONRKEAKIESGYEIPFT 634  
Db 393 SGT---TVSIGGIDFFFDLVFLFSLLEQGTGFNVNENPKILTLDNQEALEIKVQVVVFP 449  
QY 635 VTSI---ANGGSSTNELKKAVLGLTVPNTIPDG-QIIMTVKINKDS---PACQASGNQ 687  
Db 450 -TGKYDVGNGPIITYDKVGLGELKITPR-TGETVRLVIELKQEIITGYLTNEVSGVNY 508  
QY 688 TILCISTKNLNTQAMVNGGTLIVGGIYEEDNGNTLTKVPLGDIPIVGNLFTKTRGKKT 747  
Db 509 TVPITSNRELNSDVVWENGRTWVGGLISRSKLSKSTKIPGLGDIPIVWGLFRYDRDEKD 568  
QY 748 RRELLIFITPRIM 760  
Db 569 KTSLFIFLTPYVI 581

RESULT 29  
AD1935  
General secretion pathway protein D [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
R:Accession: AD1935  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigun  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AD1935  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-823 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA072988.1; PID:G17130377; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr1031  
Query Match 8.1%; Score 311; DB 2; Length 823;  
Best Local Similarity 19.3%; Pred. No. 4.1e-10;  
Matches 178; Conservative 138; Mismatches 310; Indels 296; Gaps 38;  
QY 10 SCLFVATRAAF-----QTASAGNITDIKVSSLENKQKIVKVSFDEKEIVNFTGFTSSPA 62  
Db 7 NGLVLGAARFVFLAAQPVSAQTQTITEVKLN--PONGGLSVLTKTSAGSRPQVFTTKRGK 64  
QY 63 RIALDFEOTGISMDOQVLEYADPULSKISAQNSRARLVNLNKKP--GQNTVEV---RG 117  
Db 65 ALVADIINTQLRLPQ-----GNNFRQ-----DKPAAGIASVEVWQJDA 102  
QY 118 NKWVIFINESDDTVSAPARPAKAPAKAQOQGRVYQVRSIRIQTLYEGKTTAAAPF 177  
Db 103 NSIRIVITGDS-----APTSP---VIRQOQGITLSTPTGTITASPV 144  
QY 178 TESVSVSAPESPAPKQQAASAKQQTAAAPAKQQAQQAAPAKQTNIDFRKQKNA 237  
Db 145 T-----PRPVSFTTPASTPATG-----VVPNVLPNPQVITD-GKPAQPA 185  
QY 238 GIIEALALGAFAGQPDISQOHDHIIIVTLKNHTLPTTLQRLSDVADFKTPQVQVTLKRLND 297  
Db 186 G-----PQGP-LSQ-----APPELPRAV-----APPVGDIAISATD-- 215  
QY 298 TQLIITTAGNMELVNKSAAPGYFTFQVLPKKNLESQGVNNAKPTFTGKISLDFQDVEI 357  
Db 216 -----ASP-----STIDLTQSERVPR-----LVLRDAPV 239  
QY 358 RTILOILAKESGMNIV-ASDSVNG-----KMTLSLKDVDPWDQALDLMQARNLD 405  
Db 240 REVLSLLAFANLNLAYISDGVGTPAAGGQGISQTLSDIENEPEVQDVFNVLRLSGLE 299  
QY 406 MRQOQGNVNIAPR-----DELLAKDAFLQAEKDIA-----DLGALYSQNFQKLYKNYEE 455  
Db 300 ANRSNRTIFVGPKLPNSTRDVVMRLNRLNVGNVVALNVLGLGAETAVSRERQVTVNA 359  
QY 456 F---RSIURLDNADTT-----GNRTLVSGRGSVLIDPATNLTIVTDRSVIE 500  
Db 360 VPVGTGVAPITQTTTETRVETQRFQDSNPLRLGL-QALGDBERTNSLTIGPFPKIVE 418  
QY 501 KFRKLIIDELVPAQOQWTEARIVE-----AAGDGSFSDL---GV 535  
Db 419 MAMNQLTQJDIRRQVNVNKKIIDVNLNTQDNNASLSFGIGNNYFSSDGGGAASLNIGL 478  
QY 536 KFGATGKKK-----LKNQ-----TSA 551  
Db 479 NPPSAGNVNNSLNPVTNTPYSSGNTFLDLTQSVGPVPTGVDTRNTINPNFGGGSITET 538  
QY 552 FGMGVN-----SGFGDDKKWGAETKINL-----PITAAANSISLVRAS 590  
Db 539 QGQDLNFTVORRAGISGDP---FETGITAVTQGTNPRIIRTTTDTTGTGAVTIVTGT 595



```

Qy 591 SCALN-----L ELSASELSKTKTLANPRVLTQNRKEAKIESGVEIPFTVT 636
Db 596 RGSITAALEPSLFQYKRLANLQAVTNGNAKILTDPTLIVQEGQTANVKLTQEVVGNIT 655
Qy 637 SIANGSSNTPE---L KKAVLGLTIV---PNITPDGQIIMT---VKINKDSPAQCCASN 686
Db 656 SETTRGDGASTQTVAETKTDVGLTLAVKIDRI DDNGFEVSLVAPVKAQCSSAVINVSNG 715
Qy 687 -OTILCISIKNLNT-QAMVENGGLTIVGGIIVEEDNGNTLTQVPLLCGDIPIVGNLTKTRGK 744
Db 716 SQTIFLVSRSLNSGMILRGGQTLISGLIIQQDRTSVSKIPILGDIPLIGSLFRKTNR 775
Qy 745 KTDRELLIFITPRIMGTAGNS 766
Db 776 TNERREVILLTPQWDDSQNS 797

RESULT 30
S08084
gene IV protein - phase I2-2
C;Species: phase I2-2
C;date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 20-Sep-1999
C;Accession: S08084
R;Schoenmakers, H.F.P.M.; Yu, M.; Konings, R.N.H.
submitted to the EMBL Data Library, February 1989
A;Reference number: S08084
A;Accession: S08084
A;Molecule type: DNA
A;Residues: 1-428 <SCH>
A;Cross-references: EMBL:X14336; NID:gl4920; PIDN:CAA32520.1; PID:gl4922
C;Genetics:
A;Gene: IV
C;Superfamily: filamentous phage gene IV protein
C;Keywords: phage maturation

```

```
Query Match      8.0%; Score 309; DB 2; Length 428;  
Best Local Similarity 22.8%; Pred. No. 2e+10;  
Matches 104; Conservative 85; Mismatches 179; Indels 88; Gaps 14;
```

QY 336 VNNAPKFTGCKISLDFODVEIRTIQLILAKESGMNIVASDYNKGWKTLSLKOV---PWD 392  
DB :|||::-----VRSFVOWYSKKTKGSVINVPDVXGNITVFVNADVVNNANID 70

QY 393 QALDLVQARNLMVRQQGVNIINRAPDELLAKDKAFLOAEKDIALDALYS-----443  
DB ::||::-----

Db 71 DFFKSVLNAGLA-VVAAGNPAAVSTPLTLKLASOPSNEETYDDSD -GVAYEAVPQSAAPA 128

QY 444 -----QNOLKYKNVEEFERS--IIRLDNADITTGNRTLVSGRGSVLIIDPARNTLTIVTD 494  
DB :|||::-----GGNVVDVIPGNNSLVVSIG 177

Db 129 VPADLTRVNF----NVRTRVSSDVDLEAXIFVDSN-----CGENVVDVIPGNNSLVVSIG 177

QY 495 TRSIEKRKLIDELOVPAQVMIEARIVE--AADGEFRDLGVKFVGATGKKKLKNDTSASF 552

Db 178 SAQMFPALSDFITSIIDVAREQVLIIQSLMFEETSVNGVDLSFALLAASGGKVAGGFNTSAL 237

QY 553 GWGVNSFGGDGWGAETHKINLPITAANSISILVRATSSCALNLIELSASELSKTTKAN 612

Db 238 G-----TALSTAGSGFI---FNIGNILAISOAQVOSDNSSKVIEST 274

QY 613 PRLVTQRKEAKIESGYEIPTVTSIANGSSNT-----ELKAVLGILTVENTIPDGQ 667

Db 275 PRILTOSGGYGYSVGONPVFPVKGTVGEAASNPNPQTIERRDVGSLKVTPEWMGNGQ 334

QY 668 IIMTVKINKDSPAACASGNOTI---LCISTPKNLNQAMVENGTFLVGGYEEEDNGNTLT 724

Db 335 LVLTIDTKADS-----LSNOAIASDIITNRQCIQTVOIKOGDTLLGGSISSNQFESDR 389

QY 725 KVELLGDGPVIGNLFKTRGXKTDREELLFITPRIM 760

Db 390 SVFMFKGIPIGLWFRRSHSDSKDDRMTFWILLTAHVI 425

RESULT 31

AF0100  
 general secretion pathway protein D [imported] - *Yersinia pestis* (strain CO2)  
 C:Species: *Yersinia pestis*  
 C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
 C:Accession: AF0100  
 P:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, G.  
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.  
 il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel  
 Nature 413, 523-527, 2001  
 A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.  
 A:Reference number: AB0001; MIMD:21470413; PMID:11586360  
 A:Accession: AF0100  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-640 <KUR>  
 A:Cross-references: GB:AL590842; PIDN:CAC89665.1; PID:gi5978892; GSPDB:GNO0175  
 C:Gene-ids:  
 A:Gene: YPO0816

```

Query Match      8.0%; Score 307.5; DB 2; Length 640;
Best Local Similarity 20.1%; Pred. No. 4.5e-10;
Matches 117; Conservative 97; Mismatches 182; Indels 187; Gaps 16;

Qy      340 PKTFTGRKISLQFDQVEIRTIQLILAKESGMNIVASDSVNG-----KWT 383
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      23 PLDIISANFSVSFKVDIKEFINSVSKNINKTIIIDPTVQGLISIRSIVENLKDXTYYQLF 82
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      384 LSLKDV-----PWOQALDLWQARNL----- 404
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      83 LNVLDVGYAAETMPHNVLKVTSSKRAKGVVAPLPKEGVTFDGDLEINRVLPRVISAKK 142
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      405 -----DMRQOQGNIVNIAPIRELLAKDKAFL--QAEKDIADLG-----ALYSQNF 446
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      143 ITPLLRQLNDNTESGSIINYPDSNILLITGRAAVVNHLSIVTLDQAGDNEIELYKLN 202
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      447 QL-----KYKNVEEPRSIIRLDNADTTGNENTLVSG----- 477
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      203 ATAADVVKIWEAINPINLKEVSIQVVIADERTN-SILISGDTYTRKKSILMIKKLD 261
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      478 -----RGSVLID-----PATNT 489
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      262 KRQSDGNTKVVVMKYAQASKLLDLVNGISEGFHNEKKTQSNQWNPVAKAYDQNTNA 321
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      490 LIVTDTSRVIEFKRLIDELDPVQOVMIARIVEAADGFSDLCVKGFGATGKKLKNDT 549
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      322 LVITADPPMKALGEVIEKLDIRRAQVLVEALIVETQNGEGINLVKW----ENKRSDDI 377
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      550 SAFGHWGVS-FFGGDDKKGAEKINLPITAAANSISILVRAISSGAGNALNELSASELSKTK 608
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      378 NFI---KNSDGLUNNGWGIAITF-----GLTAGFYKGNWDVLLSALSTNTWNN 424
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      609 TLANPRVITQNRKEAKISGVEIPFTVTSIANGSGSSTNTLKKAVLG--LTVTNPITPDG 666
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      425 ILATPSIVTLDNNEAEFNVQVEPVLISQTQTITTDKVNYSISQSIGVWLKVKPQINKGD 484
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      667 QLIIMTVKINKDSPACASGNQTLICISKNL-----NTQAVENGSGTLIVGGIVEE 717
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      485 SVLLIRQEVSSIASDST-----VNTNLSGVFNKRVNNAVNLKGETVVGGLUDK 537
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      718 DNGNTLTKVPLIGDIPVIGNLFTKGTGKTDRELLIFITPRIM 760
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
      Qy      538 KSGSTVKNVPPFLGDLPLGWLFRQTKKEVKSNNLIFIKPTIL 580
      Db      : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 32

C41843

pefD protein - *Xanthomonas campestris*  
C:Species: *Xanthomonas campestris*  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Oct-1999  
C:Accession: C41843; S27745  
R:Ru, N.T.; Hung, M.N.; Chiou, S.J.; Tang, F.; Chiang, D.C.; Huang, H.Y.; Wu, C.Y.  
J. Bacteriol. 174, 2679-2687, 1992  
A:Title: Cloning and characterization of a gene required for the secretion of extracellular

A:Reference number: A41843; MUID:92210513; PMID:1313415

A:Accession: C41843  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-759 <HUI>  
A:Cross-references: EMBL:M81648; NID:g155398; PIDN:AAA27615.1; PID:g155391  
A:Experimental source: pv. campestris  
A:Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBIN:92688)  
C:Genetics:  
A:Gene: pefD

Query Match 7.9%; Score 305.5; DB 2; Length 759;  
Best Local Similarity 19.9%; Pred. No. 7.6e-10; Indels 253; Gaps 32;  
Matches 162; Conservative 123; Mismatches 275; Indels 253; Gaps 32;  
Matches 162; Conservative 123; Mismatches 275; Indels 253; Gaps 32;

QY 84 DPLLSKISAAQSSRARLVNLNKPQYNTVEGRKKNVIFINESDDTVSAPARPAVK--- 140  
DB 37 DPVAGAGATQTAEQR-----ADGNASAKPTPVIRGGS 70  
QY 141 -----AAPAAKQOGCCT---VYVRSIRIOTLYP-----GKTTAAAPFTESVVS 183  
DB 71 GTMINQSAAPSPPTLGMASGSSATPFEGESVQAVVKAAILGDMGQNVIAVPGVQGIT 130  
QY 184 VSA--PFSPAK-----QAAAASAKQOTAAAPAKQOTAAAPAKQAAAAPAKQTN 228  
DB 131 LATPNVSPAQAALNLLMLVGNWARMVFGGRYINVPADQALAGTVAPSTASPAARGF 190  
QY 229 DFKDGNAGIIEAALGAGQPDISQOH---DHIIVTLKHTLPT-----TLQRS 276  
DB 191 EVR-----WVPLKYISASEMKKVLPEYARPNAIVGTDASRNVIILGTRAELENYLKT 243  
QY 277 LDVADFK-----TPVQKVTLKLNNDTQIITAGNWLNVNSAAPGYFTFOVLPK 327  
DB 244 VQIFDVLWLSGMSVGFPIQSGAKESLADLEKVF-----EQSKTPSGMPFEMP- 294  
QY 328 KNLSEGGVNAKPTFTGRKISLDFQDVEIRTILOILAKESGNIVASVNGKNTLSLK 387  
DB 295 ---LENA---NAVLIPTQPRYLD---QIQWLDRI-----DSAGGVLFYSY 333  
QY 388 DVFPDQALDVMQARNL-----DMRQOCNINVI-----APRDELLAKDAFL 429  
DB 334 ELRYIKAKDLADRLSEVFGGRGNGSGPSLVFGVVMNLGNSGGADDESLSGSSGAT 393  
QY 430 QAKEDIALDGLALYSONFQLYKKNVEPRSLRLDNADTTGNRNTLVSGRGSVLIDFA--- 486  
DB 394 G-----GDIGTSGNSGQ-----SGTSGSFGG-SSGSGMLQLPFTNQ 430  
QY 487 -----TNTLIVTDRSVIEKPKKLIDELDVPAQVQVMEARIVEAA 526  
DB 431 NGSVTLVEGDKGVSAVAETNLLVTSIAQAWKSIRDVIEKLDVMPQVHIEAQIAEVT 490  
QY 527 DGFSDRLGVKFGATGKKLKNDSIAFGWVN-----SGFGGDKKWAETKIN 573  
DB 491 -----LTGRLO-----YGVNWPENAVTTPSNADGSGGP-----N 520  
QY 574 LPITAAANSI--SLVRAISGALNLE-----LSASELSKTKTLANPVLTONRKE 622  
DB 521 LP-SAAGRGIGDVGSGVTSNGVATFLGKNAALISALDOVTNRLLOTFSVFRVNAE 579  
QY 623 AKIESGYEIPFTVTSIANG-----GSNTNELKAVGLTVPNTPDG-----QIIMTVKI 674  
DB 580 ATLNVGSRIPINSTINTGLGSDSFSVQVIDTGVILKVRPRVTKDMVFLDIVQEVST 639  
QY 675 NKDSPAQCSAQNTIL-----C---ISTKNLTOAWVNGGTLIVGGIYEEDNGNLTAKVP 727  
DB 640 PGARPACTAAATTNVNSAACNVDIRRVKTEAAVQNGDIIIMLAGLIDDDSTTDSGNIP 699  
QY 728 LIGDIPVIGNLFTKRGKTDRELLFITPRIM 760  
DB 700 FLKSLPVVVALFGRKTONSDRREVILITPSIV 732

## RESULT 33

F71486  
probable yopC/gen secretion protein D - Chlamydia trachomatis (serotype D, strain UM3,  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
C:Accession: F71486  
R:Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t;  
A:Reference number: A71570; MUID:9900809; PMID:9784136  
A:Accession: F71486  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-921 <ARN>  
A:Cross-references: GB:AE001337; GB:AE001273; NID:g3329113; PIDN:AA66269.1; PID:g332  
A:Experimental source: serotype D, strain UW-3/Cx  
C:Genetics:  
A:Gene: yscC

Query Match 7.9%; Score 303; DB 2; Length 921;

Best Local Similarity 20.3%; Pred. No. 1.4e-09;  
Matches 204; Conservative 139; Mismatches 268; Indels 396; Gaps 38;

QY 2 NTKLTKIISGLFVATAAFQTASAGNITDI--KYSSLNKKQKIVKVPDKEIVNPTGFVTS 59  
DB 18 NKKLGLSALVLDLALLSVNSRSSEGLIGQSALFN----- 55  
QY 60 SPARIALDFEQTGISMDDQVLEYADPLLSKISAAQSSRARLVNLNKPQYNTVEGRNK 119  
DB 56 -----YHET-----EQOI-----AACPKNIAMNLAACKSPG----- 81  
QY 120 VMIINESDDTVSA--PAP-APKAAAPAP---AKOQGRVTYQVRSIRIOTLYPGKTT 172  
DB 82 -----SKPTVGASFPSPQVSVKAAAPAKPQTPVAQTRHFKKSHQIFS----- 122  
QY 173 AAAPTESVSVSAPFS---PAKQAAAASAKQOTAAAPAKQOTA---APAKQAAAAPAKQ 225  
DB 123 ---PNTQSPQVKNPEERRRPLESKYLGAVKQAAA-AKEKKALEQEVSKQEEBASKLWE 179  
C/ 226 TNIDPRKGNAGIIEAALGAGQPDISQOHDLIIVTLKNTLPTTLQRLSDVADPKTP 285  
DB 180 EKQSVARAVN-----AINFSVRKQIEEQK---TISPNQNDOTLPRKDDQTSSEP 228  
QY 286 V-----OKVTLKRLNNDTQIITAGNWLNVNSAAPGYFTFQVLPKQKHL 332  
DB 229 VIQTQDCSQDQEEKKVLERLN-----IVASD--S 377  
QY 333 SGVNNAPKPTFTGRKISLDFQDVEIRTILOILAKESGNV----- 377  
DB 257 CODLKEVEYT-----VNFEDISILELQFVSKISGTFVDSNDLQFNVTIVSHDFTS 309  
QY 378 VNGKMTLSLKDVPWQALDLMQARNLDMRQOCNTV----- 413  
DB 310 VDDLATILLO-----VLKXHDLVKVEQGNVLYRNPKLSLSTVVDGSAKDTG 359  
QY 414 -----NIAPRELL-- 422  
DB 360 EAVVTRFRLYSVSPSAAGIIOPLLSDHAIISASESTRHIIIVSDIAGNIEKVELLOA 419  
QY 423 -----AKDKAF----- 428  
DB 420 LDPGTAIDMSEYDVGFANPAALVSYQDVLGAMAEBAFQIFQGTNKFIVISSPRLT 479  
QY 429 -----LQAEKDI-----ADLGALYSON-----POLKYKN-----VE 454  
DB 480 AKTIQLLESLDIPMAHTLDDVTSPAAALGSSGAANPKSLRFFMYKLYKQNGAAATAQIQ 539  
QY 455 EFSIILRLNADTGNRNTLVSGRGSVLIDPATNTLIVTDRSVIEKPKKLIDELDVPAQ 514  
DB 540 DIGYNLYVTTAMDEDFINTL-----NSIQWLPVNSIVVIGNQANVDKVVSLNGLDLPPK 595  
QY 515 QVMIEARIVEAADGFSRDLGVKFGATGKKLKNDSIAFGWVNSGFGGDDKGAETKINL 574

Db 596 QYVIEVLLETLEKSWDFGVMAALGDDEGK---VAVASGLLSNTGLTDPRLNQ---SL 649

QY 575 PITAAANSISL-----VRAIS-----SGALNLE-----LSASESL 605

Db 650 VPAPNPGNISLPPOGLAGISDMWYSSAFGLIIGNVLSHNGKSVLTGLGLLSALDQDG 709

QY 606 KYKTLANPRVLTONRKEAKIESYEIPFVTS--TANGSSST-NTELKKAVALGLVTNMI 662

Db 710 DTTVLNPRMAQDTCQAFVFGQTIPFOTTSVIOETSGVTONIEYEDIGVNLVVTSTI 769

QY 663 TPDGQIIMTVKINKDSPACASGNQITLCISTKNL-NTQAMVENGTLIVGGIYBEDNGN 721

Db 770 APNN--VVTLOI-EQITSELHSAQGVLTPTVDTKFAATRLQVPDGCFLVMSGHIRDKLT 826

QY 722 TLTKVPLLDGIPVIGNLFTKRGKKTDRRELLIFITPRIM-----GTA 763

Db 827 IVSGVPLLSPLIKGLFSRQDQKRNIMIFPKPKVISSFEFGTA 873

RESULT 34

C83411

secretion protein Xqha PA1868 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: C83411

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, S.; Lory, S.; Olson, M.V.

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: C83411

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-776 <STO>

A:Cross-references: GB:AE004613; GB:AE004091; NID:g9947856; PIDN:AAG05257.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: xqha; PA1868

Query Match 7.8%; Score 300; DB 2; Length 776;

Best Local Similarity 21.4%; Pred. No. 1.6e-09;

Matches 122; Conservative 97; Mismatches 180; Indels 170; Gaps 16;

QY 349 SLDFQDVEITLIQILAKESGMNIVASDSVNGKMT-----LSLKDVDPQALDLMQAR 402

Db 47 TINMKDAEIGDFIEQVSSISGQTFVDPYKGRVTVVSQARSLAEV--YQLFLSVLATH 104

QY 403 NLMROQGNIVNTAPRDELLAKKAFLOAEKDIAAD----- 437

Db 105 GYAVLPQGDQARIVPME--ARQDA---AQKTVRDGPGSLTEVVAQQTSAELIPMIR 159

QY 438 -----LQALYSQNQL---KYKNVEEPRSIIR----- 461

Db 160 PLVPAHGLAAVPSANALIVSDRRSNRIETAEIVRSLDRAGEHDYGIYDMRHAWAEIAE 219

QY 462 -LDN-----ADTTGNRNTLV----- 475

Db 220 VLRSVTPAAGKSAATVOVLADSRNRLVLLGPPQARALLRLAQSLDVPSSRSANSRVI 279

QY 476 -----SGRGS-----VLIDPATNTLIVDTSTRVIEKF 502

Db 280 RLRHGDAKTLAATLIGESLHGERGQDGRSGKRGILLVRADESINALVILADPEDVGLL 339

QY 503 RKLIDEUDVPAQVMIEARIVEAADGFSRDLGVKFGATGKKLKNDTSAFGMG-VNSGFG 561

Db 340 EDIVRQDLVPAQQLLVAAALVELSGEIGDALGVQWA-----LRSGHVAGGAFDAGSL 393

QY 562 GDDKWAETKXINLPITAAANSISLVRAISSGALNLELSASESLSKTKTLANPRVLTONRKE 621

Db 394 IGTLLGALQAGKPPAELPDGAI VGLGRDFGALVLTALSRN---SRNLLSTPSELLTLDNQ 450

QY 622 EAKIESGYEIPF---TWTSTANGSSSINTELKKAVALGLT--VTPNITPDGQIIMTV--KI 674

Db 451 KAEILVQNVFQTSYTTTSAGSSNFFTVRDXDVGTLKVTPHIGEDRMRLIEIQEI 510

QY 675 NKDSPACASGNQITLCISTKNLNTQAMVENGTLIVGGIYBEDNGNTLTKVPLLDGIPV 734

Db 511 SSIAPTATLAAKAVDLVTNKRISIKSTVLADDDQGVILGGLIQDDLQSRSDSRVPLLDGIPG 570

QY 735 IGNLFTKRGKKTDRRELLIFITPRIMGTA 763

Db 571 VGLFRSSRSTRVKNRLMWFLRPSIVRDA 599

RESULT 35

Z48PFI

gene IV protein - phage f1

C:Species: phage f1

C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 23-Jul-1999

C:Accession: C04268; A04268

R:Beck, B.; Zink, B.

Gene 16, 35-58, 1981

A:Title: Nucleotide sequence and genome organisation of filamentous bacteriophages f1

A:Reference number: A91490; MUID:82211801; PMID:6282703

A:Accession: C04268

A:Molecule type: DNA

A:Residues: 1-426 <BEC>

A:Cross-references: GB:V00606; GB:J02449; GB:M10881; NID:gl4974; PIDN:CAA23875.1; PID

C:Comment: The exact function of this protein is unknown although it may be involved

C:Genetics:

A:Gene: IV

C:Superfamily: filamentous phage gene IV protein

C:Keywords: phage maturation

Query Match 7.8%; Score 298.5; DB 1; Length 426;

Best Local Similarity 20.9%; Pred. No. 8.1e-10;

Matches 91; Conservative 96; Mismatches 189; Indels 59; Gaps 11;

QY 350 LDFQDVEITLIQILAKESGMNIVASDSVNGKMTLSLKDVDPQALDL---VMOARNLDM 406

Db 24 IEMNSSLRQDFVTWYSKQTESIVSPDKGTIVYSDVRPENLDRPFFISVLRANNFDM 83

QY 407 RQQGNIVNIAPR-----DELLAKDKAFLOAEKDIAADLALY-----SQNFOLKY 450

Db 84 --VGSIPSIIOKYNPNNDYIDELPSSDNQ--EYDDNSAPSGGFFVPQNDVNTQTFINN 139

QY 451 KNVEEPRSIIRLIDNADTTGNRNTLVSGRGSVLIDPATNTLIVTSTRVIEKRLIDELD 510

Db 140 VRAKDLIRVLEIFVKSNTSKSNVLSVDGS-----NLLVVSAPKDLNDLPQLSTVD 192

QY 511 VPAQQVMIEARIVEAADGFSRDLGVKFGATGKKLKNDTSAFGMGVNSGFGGDDKWAET 570

Db 193 LPTDQILLEGILFEVQQGDALDFFSFAAG-----SQRGTVAGGVNT-----D 233

QY 571 KINLPITAAANSISLVRAISSGALNLELSASESLSKTKTLANPRVLTONRKEAKIESGYE 630

Db 234 RLTSVLSSAGGSFGI---FNGDVLGLSVRLAKTNHSHKILSVPRILTLGSGKGSISVGQN 290

QY 631 IPFTVTSIANGSSSTN---TELKKAVALGLTVPNITPDGQIIMTVKINKDSPACQACSG 685

Db 291 VPFTIGRTVGTGSANVNFQTVERNQVGISMSVFPVAMAGGNIVLDTSKADSLSSSTQA 350

QY 686 NOTILCISTKNLNTQAMVENGTLIVGGIYBEDNGNTLTKVPLLDGIPVIGNLFTKRGK 745

Db 351 SDVI--TNQSGIATTVNLRDGGTLLGLLTDYKNTSQDSGVFFLSKIPILIGLILFSSRSDS 408

QY 746 TDRRELLIFITPRIM 760

Db 409 NEESTLYLVIVKATIV 423

RESULT 36

G82326

MSHA biogenesis protein MshL VC0402 [imported] - Vibrio cholerae (strain N16961 serog

C:Species: Vibrio cholerae

C.Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C.Accession: G82326  
C.R#Accession: G82326  
R.Heldberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
R. Charlestone, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.  
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A.Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
A.Reference number: A82035; MUID:20406833; PMID:10952301  
A.Accession: G82326  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-559 <HEI>  
A.Cross-references: GB:AE004128; GB:AE003852; NID:9654815; PIDN:BAF93575.1; GSPDB:GN001  
A.Experimental source: serogroup O1; strain N16961; biotype El Tor  
C.Genetics  
A.Gene: VC0402  
A.Map position: 1

|    | Query Match           | 7.7%;   | Score 296;         | DB 2;       | Length 559; |
|----|-----------------------|---|--------------------|-------------|-------------|
|    | Best Local Similarity | 21.2%;  | Pred. No. 1.7e-09; |             |             |
|    | Matches 107;          | Conservative 103;   | Mismatches 180;    | Indels 114; | Gaps 18;    |
| QY | 340                   | PKTFTGRKISLDFQDVEIRTILOILAKESGMNIVASDSVNGKXVTLSLKDVDPDQALDLYM | 399                |             |             |
| DB | 65                    | PKTL--QRPRIAQEDVEAKFAFFASLVQGTYSAAIHPAVTTGRITLNLDTVTDLGALGVVR | 122                |             |             |
| QY | 400                   | QARNLDMRQOGNIVNAP---RDELLAKAKAFQAEKDIADLGALY-----SONFQL       | 448                |             |             |
| DB | 123                   | DLYGFEVVEGKGVQVTPAGLRVTTPVD--YLQFKTGRSLTSTITGITINTDTNNSNS     | 180                |             |             |
| QY | 449                   | KYNVVEEFRSLRDLNADTTGNRN-----TLVSGRG--                         | 479                |             |             |
| DB | 181                   | SSSSSSSISSNSSDSSSSNSNSNRSDARGTETIETTNESDFWPLEEKAVAQLLGGSGGQ   | 240                |             |             |
| QY | 480                   | SVLIDPATNTLIVTDRSVTEKFRKLDLDELVAQ----QVMIEARIVEA--ADGFSRDL    | 533                |             |             |
| DB | 241                   | TVIVNPAQVLTL---RAYDEIRVQNEFLGISQORMHRQVILEAKILEVTLSDGYCQGI    | 297                |             |             |
| QY | 534                   | GVRFGATGKKLKNDTSAF-----GWGVNMGFGGDDKRWGAETKINLPITAAANSISLVR-  | 587                |             |             |
| DB | 298                   | -----NWSKAFSSNGANYKIGSGITQDSNGN-----PITSVLPLGLDAIGN           | 338                |             |             |
| QY | 588                   | -----AISSGALNLELSASELSKTKTLANPVLTONRKEAKIESGVEIPTVTSIA        | 639                |             |             |
| DB | 339                   | LLGQSNVIVSSGSFDAVIFPMATQGDNLVLSRPVTSASNNQKAVIKVGTID-EYVVTDL   | 397                |             |             |
| QY | 640                   | N-GGSSNTTELKKA-----LGLVTNNTPDGQIIMTVTKINKDSPPAQCSAGNOT--      | 688                |             |             |
| DB | 398                   | SVVGTGDNAQASPDITLTPPFGSIGSLDVTPOIDQGNVLLHV-----HPAVIEVEQQT    | 452                |             |             |
| QY | 689                   | -----ILCISIKVLNTQAMVNGGTLIVGGIVEEDNGNTLTKVPLLGLDIPVGN         | 737                |             |             |
| DB | 453                   | ILYKSEELPLAKSSRESVSTRAKDDGVVIGGLMKSNVTVDQVSKVPLGLDVPALGH      | 512                |             |             |
| QY | 738                   | LFKTRGKKTDRRELLIFITPRIMG                                      | 761                |             |             |
| DB | 513                   | LFRNTTKLTOKTELVIILKPTWVG                                      | 536                |             |             |

RESULT 37  
I39547  
S-protein secretion D - Aeromonas hydrophila  
C:Species: Aeromonas hydrophila  
C:Date: 19-Jul-1996 #sequence\_revision 19-Jul-1996 #text\_change 08-Oct-1999  
C:Accession: I39547  
R:Thomas, S.R.; Trust, T.J.  
J. Bacteriol. 177, 3932-3939, 1995  
A:Title: A specific PnLD homolog is required for the secretion of paracrystalline surface  
A:Reference number: A57354; MUII:95332195; PMID:7608063  
A:Accession: I39547  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-737 <RES>

A; Cross-references: GB:L41682; NID:9950300; PIDN:AAA79322.1; PID:g1019923  
C:Genetics:  
A:Gene: spsd

Query Match  
Best Local Similarity 21.1%; Pred No. 3.3e-09;  
Matches 134; Conservative 95; Mismatches 217; Indels 188; Gaps 23;

QY 245 LGFAGOPDISQOHDHIIIVLKNHTLPTTLQKSLD--VADFKTPOV-KVTLKELN-----295  
DB 129 LSFMTPELQKSLDLVLRMSDPQPSVLFNTARNVLSDYGEVKLRDGLXTFNVAQTVA188  
QY 296 NDTQLIIITAGNWEVLNKSAAFGYFTFOVLPK-----327  
DB 189 NDLALPIVMSGGALPDVPMSHRP--VFQIVPMKVRSDOAAWLAEMFENSOLKIKDDRVA245  
QY 328 -----KQNLSEGGVNNAP--KTFTGRKISLDFQVEI--RTILQILAKESGM370  
DB 246 INALMLRGPVELVKQASEAISLFDOPALKSSHSLAISPYVSDAEQLGNALMKVLOSE-GY304  
QY 371 NIVASPSVNGKMTLSLKDV-----PMDQALD-----LVMQAR402  
DB 305 DVSDDSPFGVLVLMKELQRIIVFAAEPRVLISHRQVVEVLDRESSQEKVENGELFIYQVR364  
QY 403 NLDMRQOGNIIVNTAPRDELLAKDXAFLOAEKOIADLGAL-YSCNPOLKYKNVEEFSLIR461  
DB 365 N-----TQAAAIATMLGALGYGANIPATGV-----S391  
QY 462 LDNADTTGNR-----TLYS-----GRGSVLIDPATNTLIIVTDRSVIEKFR503  
DB 392 TNTVTATGSGAGSLTAASPATSPTSAPKGEQSVVVDGNNALIFKSGREWVTLR451  
QY 504 KLIDELDVPAQQVMIPARIVEAADGFSRDLGVKFGATGKKKLKNQDTSAFGWGNSFGGD563  
DB 452 PLLDELCKPVPVSMIDVLLAEVSLNDKEGLGVDWS-----NIHTDLGAK495  
QY 564 DKWGAETKINLPITAAANSISLVRATSSGALNDEL-SASESISKTKTL-----ANP613  
DB 496 DL-----IMSTAN-----GIGNSGLNLTNSAGQTRAKLNAFYENKQAVIRSSP539  
QY 614 RVLTQNRKEAKISGVEIPTVTSIANGSSSTNTELLKAV-----LGLTVTNIITPDGQ667  
DB 540 KLMVRSGEEARIEVGNBIP-VVGTGTSQSDNPDPINKTVQYRKTVGIIITKFTPOASGV598  
QY 668 IIMTVKINKDSPACOCASGNOTIL-CISTKNLTQAMVENGGLIIVGGEIIEEDNGNTLTKV726  
DB 599 VDLTISQELSEQDQDTSTSTDALPTIMNRKVNATLIRDGSVWMLAGLISSTKGEQGTGV658  
QY 727 PLLGDIEVIGNLFTKRGKXTDRELLFIITPRIM760  
DB 659 PLLGDIEVIGSLFKPSKNSONRTLELVMMIIPYVL692

RESULT 38  
 ZABPFD  
 gene IV protein - phage fd (strain 478, Heidelberg)  
 C:Species: phage fd  
 C>Date: 30-Sep-1980 #sequence\_revision 30-Sep-1980 #text\_change 23-Jul-1989  
 C:Accession: A04268  
 R:Beck, E.; Sommer, R.; Auerswald, E.A.; Kurz, C.; Zink, B.; Osterburg, G.; Schaller, N.  
 Nucleic Acids Res. 5, 4495-4503, 1978  
 A:Title: Nucleotide sequence of bacteriophage fd DNA.  
 A:Reference number: A93690; MUID:79136480; PMID:745987  
 A:Accession: A04268  
 A:Molecule type: DNA  
 A:Residues: 1-426 <BEC>  
 A:Cross-references: GB:J02451; GB:M10731; GB:M10767; GB:M21666; GB:M21667  
 C:Comment: The exact function of this protein is unknown although it may be involved  
 C:Genetics:  
 A:Gene: IV  
 C:Superfamily: filamentous phage gene IV protein  
 C:Keywords: phage maturation

```
Query Match 7.6%; Score 292.5; DB 1; Length 426;
Best Local Similarity 20.7%; Pred. No. 1.8e-09;
Matches 93; Conservative 96; Mismatches 185; Indels 75; Gaps 12;

QY 336 VNNAPKFTGRKISLDQFQVEIRTIQILAKESGMNIVASDSVNGKMTLSKDKVPMQDAL 395
D 26 MNNSP-----LRDFVTWYSKQTGESVIVSPDKGTVTVYSSDVKPENLR 69
QY 396 DL---VMOARNLDMRQGNIVNIAPR-----DELLAKDKAFLQAEKDIALGALY- 442
D 70 NFFISVLRANNFDM--VGSIPSIIQKYNPNQDYIDELPSSD--IQEYDONSAPSGGFV 125
QY 443 -----SONFOLKYKNVVEFRSILRLDNADTTGNRNTLVSGRGSVLIDPATNTLIVTDTR 496
D 126 PONDNVTTQFKINNVRADLRVVELFVKSNVSVSDGS-----NLLVVSAPK 178
QY 497 SVIEKFRKLIDELVPAQQVMIEARIVFAAOGFRDLGVKFGATGKKLKNDTSAFGWV 556
D 179 DILDNLQFLSTVDLPTDQIILIEGLIFEVQQGDALDFSAAG-----SQRGTVAGGV 230
QY 557 NSGFGDDKWAETKINLPITAAANSISLVRAISSGALNLELSASELSKTKTLANPRVL 616
D 231 NT-----DRLTSVLASAGSFGI---FNGDVLGLSVRAKTNHSHKILSVPRIL 276
QY 617 TQNRKEAKIESGYEIPFTVTSIANGSGSTN-----TELKKAVLGLTVPNITPDGQIIMT 671
D 277 TLSGQKGSISVGNVPFITGRVTGESANVNNPFQTVERNQVIGISMVFPVAMAGGNIVLD 336
QY 672 VKINKDSPAACASGNQITLCISTKNLNTQAMVENGTLIVGGIYEDNGNTLTQVPLIGD 731
D 337 ITSKADLSUSSSTQASDVI--TNQRSIATTVNLRDQGTLLGLLTDYKNTSQDSGVFFLSK 394
QY 732 IPVIGNLFRKTRGKKTDRRELIFITPRIM 760
D 395 IPLIGLFSRSDSNEESTLYVLVKATIV 423

RESULT 39
Z4BPM3
gene IV protein - phage M13
C:Species: phage M13
A:Note: host Escherichia coli
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Nov-1996
C:Accession: B04268
R:van Wazerbeek, P.M.G.F.; Hulsebos, T.J.M.; Schoenmakers, J.G.G.
Gene 11, 129-148, 1980
A:Title: Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: comparison
A:Reference number: A91470; MUID:81067903; PMID:6254849
A:Accession: B04268
A:Molecule type: DNA
A:Residues: 1-426 <VAN>
C:Comment: The exact function of this protein is unknown although it may be involved in
C:Genetics:
A:Gene: IV
C:Superfamily: filamentous phage gene IV protein
C:Keywords: phage maturation

Query Match 7.6%; Score 292.5; DB 1; Length 426;
Best Local Similarity 21.6%; Pred. No. 1.8e-09;
Matches 97; Conservative 93; Mismatches 184; Indels 75; Gaps 13;

QY 336 VNNAPKFTGRKISLDQFQVEIRTIQILAKESGMNIVASDSVNGKMTLSKDKVPMQDAL 395
D 26 MNNSP-----LRDFVTWYSKQGESVIVSPDKGTVTVYSSDVKPENLR 69
QY 396 DL---VMOARNLDMRQGNIVNIAPR-----DELLAKDKAFLQAEKDIALGALY- 442
D 70 NFFISVLRANNFDM--VGSIPSIIQKYNPNQDYIDELPSSNQ--EYDONSAPSGGFV 125
QY 443 -----SONFOLKYKNVVEFRSILRLDNADTTGNRNTLVSGRGSVLIDPATNTLIVTDTR 496
D 126 PONDNVTTQFKINNVRADLRVVELFVKSNVSVSDGS-----NLLVVSAPK 178
QY 497 SVIEKFRKLIDELVPAQQVMIEARIVFAAOGFRDLGVKFGATGKKLKNDTSAFGWV 556
```

```
Db 179 DILDNLQFLSTVDLPTDQIILIEGLIFEVQQGDALDFSAAG-----SQRGTVAGGV 230
QY 557 NSGFGDDKWAETKINLPITAAANSISLVRAISSGALNLELSASELSKTKTLANPRVL 616
D 231 NT-----DRLTSVLASAGSFGI---FNGDVLGLSVRAKTNHSHKILSVPRIL 276
QY 617 TQNRKEAKIESGYEIPF---TVTSIANGSGSTNT-ELKKAVLGLTVPNITPDGQIIMT 671
D 277 TLSGQKGSISVGNVPFITGRVTGEIANVNNPFQTVERNQVIGISMVFPVAMAGGNIVLD 336
QY 672 VKINKDSPAACASGNQITLCISTKNLNTQAMVENGTLIVGGIYEDNGNTLTQVPLIGD 731
D 337 ITSKADLSUSSSTQASDVI--TNQRSIATTVNLRDQGTLLGLLTDYKNTSQDSGVFFLSK 394
QY 732 IPVIGNLFRKTRGKKTDRRELIFITPRIM 760
D 395 IPLIGLFSRSDSNEESTLYVLVKATIV 423

RESULT 40
Z4BPIK
gene IV protein - phage Ike
C:Species: phage Ike
A:Note: host Escherichia coli
C:Date: 28-May-1986 #sequence_revision 28-May-1986 #text_change 28-Jul-2000
C:Accession: A04269
R:Peeters, B.P.H.; Peters, R.M.; Schoenmakers, J.G.G.; Konings, R.N.H.
J. Mol. Biol. 181, 27-39, 1985
A:Title: Nucleotide sequence and genetic organization of the genome of the N-specific
A:Reference number: A92912; MUID:85160831; PMID:3981635
A:Accession: A04269
A:Molecule type: DNA
A:Residues: 1-437 <PEE>
A:Cross-references: GB:X02750; NID:gi14942; PIDN:CAA26076.1; PID:gi14956
C:Comment: The exact function of this protein is not known, although it may be involve
C:Genetics:
A:Gene: IV
C:Superfamily: filamentous phage gene IV protein
C:Keywords: phage maturation

Query Match 7.5%; Score 290; DB 1; Length 437;
Best Local Similarity 21.4%; Pred. No. 2.6e-09;
Matches 98; Conservative 92; Mismatches 201; Indels 66; Gaps 11;

QY 322 FOVLPRKQNLSESGVNNAPKFTTGRKISLDQFQVEIRTIQILAKESGMNIVASDSVNGK 381
D 26 FNVLADFPNL-----NNAP-----VRSFVQWYSQSKNKAVVNPDKVGN 64
QY 382 MTLSLKDPVWDQALDLVMOARNLD--MRQQGNIVNIAPRDELLA-----KDKAFLQ 430
D 65 ITVENADVNOANIDDFPKSVLNANGFVLMAAGDSPGVSTPSKLPSSQQTDDDDYEDSADYV 124
QY 431 AEKDIADLGALYSONFOLKYKNVVEFRSILRLDNADTTGNRNTLV--SGRGSVLIDPATN 488
D 125 PVGDSVPVSAQPKPLDLTVRNFK---LTVRSSDVLPLAKIFVDSNGGGDVIDYFGNN 180
QY 489 TLIVTDRSVIEKFRKLIDELVPAQQVMIEARIVFAAOGFRDLGVKFGATGKKLKN 548
D 181 SLLVSGSAAIMNALADFTSIDVARQDVLIQSLMFEISLVNGVDLSFAAG-----230
QY 549 TSAPFGWVNGSGFGDDKWAETKINLPITAAANSISLVRAISSGALNLELSASELSKTK 608
D 231 -SASGDKVAGGFN-----TSALGTALTAGGSFGI---FNGNVLALSQAQVKSNSK 279
QY 609 TLANPRVLTONRKEAKIESGYEIPFTVTSIANGSGSTNT-----ELKKAVLGLTVPNIT 663
D 280 VISTPRILTQSGQGYISVGNVPFITGRVTGEANVNNPFQTVERNQVIGISMVFPVAMAGGNIVLD 339
QY 664 PDGQIIMTVKINKDSPAACASGNQITLCISTKNLNTQAMVENGTLIVGGIYEDNGNTLT 723
D 340 GNGQLVLTIDTKADSLTSQMTASDII--TNQRHMQTTVOINKDQGTLLGLLSDNTTGDG 397
```

Search completed: December 9, 2003, 10:27:47  
Job time : 28 secs

OM protein - protein search, using sw model

Run on: December 9, 2003, 10:27:53 ; Search time 48 Seconds  
(without alignments)  
2542.932 Million cell updates/sec

Title: US-09-701-271a-2  
Perfect score: 769  
Sequence: 1 MNFKLTIISGLFVATAAFO.....ELLIFITPRIMGTAGNSLRY 769

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

| Database : |       |             |        | A_Geneseq 19Jun03.* |             |  |                     |
|------------|-------|-------------|--------|---------------------|-------------|--|---------------------|
| Result No. | Score | Query Match | Length | ID                  | Description |  |                     |
| 1          | 769   | 100.0       | 769    | 21                  | AAV53895    |  | A Neisseria mening  |
| 2          | 597   | 77.6        | 769    | 21                  | AAV53896    |  | A Neisseria mening  |
| 3          | 356   | 46.3        | 769    | 21                  | AAV53897    |  | A Neisseria mening  |
| 4          | 162   | 21.1        | 723    | 24                  | ABF77693    |  | N. gonorrhoeae. ami |
| 5          | 150   | 19.5        | 720    | 23                  | ABG91062    |  | Neisseria gonorrh   |
| 6          | 9     | 1.2         | 473    | 21                  | AAV44394    |  | Moraxella catarrha  |
| 7          | 9     | 1.2         | 473    | 21                  | AAV44395    |  | Moraxella catarrha  |
| 8          | 9     | 1.2         | 473    | 21                  | AAV44396    |  | Moraxella catarrha  |
| 9          | 9     | 1.2         | 473    | 21                  | AAV44397    |  | Moraxella catarrha  |

## SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

|    |   |     |      |    |           |                    |
|----|---|-----|------|----|-----------|--------------------|
| 10 | 9 | 1.2 | 473  | 23 | AAO17573  | M catarrhalis MCA1 |
| 11 | 9 | 1.2 | 600  | 21 | AAAB14113 | Bordetella pertuss |
| 12 | 9 | 1.2 | 649  | 19 | AAW53827  | Pseudomonas XCPQ s |
| 13 | 9 | 1.2 | 649  | 21 | AAW82594  | Pseudomonas alcali |
| 14 | 9 | 1.2 | 649  | 22 | AAAB82256 | Pseudomonas alcali |
| 15 | 9 | 1.2 | 649  | 23 | AAE13660  | Pseudomonas alcali |
| 16 | 9 | 1.2 | 654  | 22 | AAU34768  | E. coli cellular p |
| 17 | 9 | 1.2 | 821  | 22 | ABG25804  | Novel human diagno |
| 18 | 9 | 1.2 | 1146 | 20 | AAW86007  | Histidine kinase C |
| 19 | 8 | 1.0 | 66   | 22 | AAW82329  | Human immune/haema |
| 20 | 8 | 1.0 | 111  | 22 | AAW53476  | Mycobacterium tube |
| 21 | 8 | 1.0 | 118  | 22 | AAW80345  | N. gonorrhoeae ami |
| 22 | 8 | 1.0 | 231  | 23 | ABG91620  | Purine/pyrimidine  |
| 23 | 8 | 1.0 | 231  | 23 | ABG91627  | Purine/pyrimidine  |
| 24 | 8 | 1.0 | 231  | 24 | ABP77597  | Pseudomonas aerugi |
| 25 | 8 | 1.0 | 776  | 22 | AAU33596  | Pseudomonas aerugi |
| 26 | 8 | 1.0 | 1024 | 23 | ABW04826  | LDL receptor bindi |
| 27 | 8 | 1.0 | 2383 | 21 | AAW15945  | E. coli proliferat |
| 28 | 7 | 0.9 | 7    | 21 | AAW62073  | PB-cadherin cell a |
| 29 | 7 | 0.9 | 10   | 13 | AAW26158  | Recombinant signal |
| 30 | 7 | 0.9 | 10   | 22 | AAW96038  | Human complementar |
| 31 | 7 | 0.9 | 10   | 22 | AAW96040  | Human complementar |
| 32 | 7 | 0.9 | 10   | 22 | AAW96222  | Human complementar |
| 33 | 7 | 0.9 | 10   | 22 | AAW96224  | Human complementar |
| 34 | 7 | 0.9 | 19   | 24 | ABP82732  | G protein-coupled  |
| 35 | 7 | 0.9 | 20   | 13 | AAW26166  | Tuberculosis antib |
| 36 | 7 | 0.9 | 29   | 22 | AAW74443  | Herpes simplex vir |
| 37 | 7 | 0.9 | 38   | 16 | AAW77779  | Goldfish derived o |
| 38 | 7 | 0.9 | 42   | 21 | AAW58555  | Rat/mouse somatost |
| 39 | 7 | 0.9 | 42   | 21 | AAW58556  | Human somatostatin |
| 40 | 7 | 0.9 | 43   | 21 | AAW16656  | Bacteriophage 192  |
| 41 | 7 | 0.9 | 47   | 21 | AAW22564  | Zea mays protein f |
| 42 | 7 | 0.9 | 50   | 20 | AAW13188  | Human secreted pro |
| 43 | 7 | 0.9 | 50   | 22 | AAU42408  | Propionibacterium  |
| 44 | 7 | 0.9 | 51   | 20 | AAW12186  | Human 5' EST secre |
| 45 | 7 | 0.9 | 51   | 20 | AAW12461  | Human 5' EST secre |
| 46 | 7 | 0.9 | 51   | 23 | AAW16016  | Human ga-15018831  |
| 47 | 7 | 0.9 | 56   | 22 | ABW10924  | Human ovarian and/ |
| 48 | 7 | 0.9 | 56   | 22 | AAW63630  | Human reproductive |
| 49 | 7 | 0.9 | 59   | 21 | AAW56875  | Zea mays protein f |
| 50 | 7 | 0.9 | 61   | 22 | AAW61382  | Propionibacterium  |
| 51 | 7 | 0.9 | 62   | 21 | AAW57329  | Zea mays protein f |
| 52 | 7 | 0.9 | 65   | 22 | AAW89250  | Human immune/haema |
| 53 | 7 | 0.9 | 70   | 21 | AAW58533  | Rat somatostatin r |
| 54 | 7 | 0.9 | 70   | 21 | AAW58534  | Mouse somatostatin |
| 55 | 7 | 0.9 | 70   | 21 | AAW58535  | Human somatostatin |
| 56 | 7 | 0.9 | 71   | 24 | ABG76371  | Translation domain |
| 57 | 7 | 0.9 | 71   | 24 | ABG76372  | Translation domain |
| 58 | 7 | 0.9 | 78   | 23 | ABP10418  | Human ORFX protein |
| 59 | 7 | 0.9 | 79   | 23 | ABP03863  | Human ORFX protein |
| 60 | 7 | 0.9 | 81   | 20 | AAW11644  | Human 5' EST secre |
| 61 | 7 | 0.9 | 81   | 22 | AAO09650  | Human polypeptide  |
| 62 | 7 | 0.9 | 81   | 23 | ABP09567  | Human ORFX protein |
| 63 | 7 | 0.9 | 82   | 22 | AAW67075  | Propionibacterium  |
| 64 | 7 | 0.9 | 94   | 22 | AAW92288  | C glutamicum prote |
| 65 | 7 | 0.9 | 94   | 23 | AAE13241  | Human linker histo |
| 66 | 7 | 0.9 | 110  | 23 | ABP05992  | Human ORFX protein |
| 67 | 7 | 0.9 | 111  | 21 | AAW45463  | Arabidopsis thalia |
| 68 | 7 | 0.9 | 116  | 21 | AAW57264  | Arabidopsis thalia |
| 69 | 7 | 0.9 | 118  | 21 | AAW11695  | Arabidopsis thalia |
| 70 | 7 | 0.9 | 129  | 20 | AAW01481  | Protein sequence c |
| 71 | 7 | 0.9 | 130  | 23 | AAW83146  | Novel secreted pro |
| 72 | 7 | 0.9 | 131  | 22 | AAW66714  | C.Glutamicum phosp |
| 73 | 7 | 0.9 | 131  | 23 | ABG80327  | C. glutamicum meta |
| 74 | 7 | 0.9 | 132  | 13 | AAW27506  | 3' rat brain somat |
| 75 | 7 | 0.9 | 137  | 21 | AAW45462  | Arabidopsis thalia |
| 76 | 7 | 0.9 | 137  | 22 | ABG24409  | Novel human diagno |
| 77 | 7 | 0.9 | 145  | 17 | AAW99570  | Wasp venom BrhX-1  |
| 78 | 7 | 0.9 | 149  | 21 | AAW11694  | Arabidopsis thalia |
| 79 | 7 | 0.9 | 149  | 23 | AAE15403  | HPV 18 E2 protein. |
| 80 | 7 | 0.9 | 159  | 21 | AAW15693  | Arabidopsis thalia |
| 81 | 7 | 0.9 | 159  | 23 | ABP04200  | Human ORFX protein |
| 82 | 7 | 0.9 | 161  | 21 | AAW57263  | Arabidopsis thalia |

|     |   |     |     |    |           |                     |     |   |     |     |    |          |                     |
|-----|---|-----|-----|----|-----------|---------------------|-----|---|-----|-----|----|----------|---------------------|
| 83  | 7 | 0.9 | 163 | 22 | ABB70935  | Drosophila melanog  | 156 | 7 | 0.9 | 295 | 21 | AAG21606 | Arabidopsis thalia  |
| 84  | 7 | 0.9 | 166 | 19 | AAW59886  | Amino acid sequenc  | 157 | 7 | 0.9 | 295 | 21 | AAG45109 | Arabidopsis thalia  |
| 85  | 7 | 0.9 | 166 | 13 | AAW49006  | Human thioredoxin   | 158 | 7 | 0.9 | 297 | 21 | AAG11324 | Arabidopsis thalia  |
| 86  | 7 | 0.9 | 166 | 23 | ABP60804  | Homo sapiens thior  | 159 | 7 | 0.9 | 304 | 21 | AAG50946 | Arabidopsis thalia  |
| 87  | 7 | 0.9 | 166 | 23 | ABP41978  | Human ovarian anti  | 160 | 7 | 0.9 | 305 | 21 | AAG05114 | Arabidopsis thalia  |
| 88  | 7 | 0.9 | 167 | 21 | AAW58948  | Breast and ovarian  | 161 | 7 | 0.9 | 307 | 21 | AAG19976 | Arabidopsis thalia  |
| 89  | 7 | 0.9 | 168 | 21 | AAG58698  | Arabidopsis thalia  | 162 | 7 | 0.9 | 307 | 21 | AAG45206 | Arabidopsis thalia  |
| 90  | 7 | 0.9 | 168 | 21 | AAG57776  | Arabidopsis thalia  | 163 | 7 | 0.9 | 308 | 22 | AGC24502 | Novel human diagno  |
| 91  | 7 | 0.9 | 172 | 21 | AAG60299  | Arabidopsis thalia  | 164 | 7 | 0.9 | 312 | 21 | ABH42809 | Human ORFX ORF2573  |
| 92  | 7 | 0.9 | 172 | 22 | AAW96695  | Human complement c  | 165 | 7 | 0.9 | 312 | 22 | AAW71830 | Human olfactory re  |
| 93  | 7 | 0.9 | 176 | 19 | AAW51216  | Insecticidal toxin  | 166 | 7 | 0.9 | 321 | 21 | AAG50945 | Arabidopsis thalia  |
| 94  | 7 | 0.9 | 182 | 21 | AAW38517  | Fragment of human   | 167 | 7 | 0.9 | 321 | 21 | AAW48754 | Rat RGH G-protein   |
| 95  | 7 | 0.9 | 185 | 21 | AAW24838  | Plant SDF encoded   | 168 | 7 | 0.9 | 322 | 15 | AAW02726 | Rat RGH G-protein   |
| 96  | 7 | 0.9 | 185 | 21 | AAW41594  | Arabidopsis thalia  | 169 | 7 | 0.9 | 322 | 17 | AAW05113 | Arabidopsis thalia  |
| 97  | 7 | 0.9 | 188 | 21 | AAW91634  | Human secreted pro  | 170 | 7 | 0.9 | 322 | 22 | AAU97002 | Human CCAAT enhanc  |
| 98  | 7 | 0.9 | 191 | 22 | ABW61016  | Drosophila melanog  | 171 | 7 | 0.9 | 322 | 22 | AAU11945 | Human CCAAT/enhanc  |
| 99  | 7 | 0.9 | 193 | 23 | ABG32385  | Mouse JFV1. Mus m   | 172 | 7 | 0.9 | 324 | 23 | AAW48227 | G-protein coupled   |
| 100 | 7 | 0.9 | 194 | 21 | AAW24837  | Plant SDF encoded   | 173 | 7 | 0.9 | 329 | 21 | AAG21605 | Arabidopsis thalia  |
| 101 | 7 | 0.9 | 194 | 21 | AAW44593  | Arabidopsis thalia  | 174 | 7 | 0.9 | 329 | 22 | ABW59237 | Drosophila melanog  |
| 102 | 7 | 0.9 | 197 | 22 | AAU33961  | Staphylococcus aur  | 175 | 7 | 0.9 | 329 | 22 | ABW59237 | Lactococcus lactis  |
| 103 | 7 | 0.9 | 198 | 22 | ABW62551  | Drosophila melanog  | 176 | 7 | 0.9 | 330 | 22 | AAW01417 | Human secreted pro  |
| 104 | 7 | 0.9 | 198 | 22 | AAU37087  | Staphylococcus aur  | 177 | 7 | 0.9 | 331 | 21 | AAW15466 | Arabidopsis thalia  |
| 105 | 7 | 0.9 | 205 | 22 | AAU16946  | Human novel secret  | 178 | 7 | 0.9 | 332 | 21 | AAW05883 | Arabidopsis thalia  |
| 106 | 7 | 0.9 | 205 | 22 | AAW89837  | C glutamicum prote  | 179 | 7 | 0.9 | 332 | 21 | AAW05883 | Arabidopsis thalia  |
| 107 | 7 | 0.9 | 214 | 21 | AAW57262  | Arabidopsis thalia  | 180 | 7 | 0.9 | 332 | 21 | AAW05883 | Arabidopsis thalia  |
| 108 | 7 | 0.9 | 217 | 22 | AAU29428  | Human G protein-co  | 181 | 7 | 0.9 | 333 | 23 | ABU51983 | Helicobacter pylor  |
| 109 | 7 | 0.9 | 217 | 23 | AGW60716  | Novel G protein co  | 182 | 7 | 0.9 | 335 | 21 | AAW19975 | Arabidopsis thalia  |
| 110 | 7 | 0.9 | 217 | 23 | ABW53804  | Lactococcus lactis  | 183 | 7 | 0.9 | 335 | 21 | AAW45205 | Arabidopsis thalia  |
| 111 | 7 | 0.9 | 218 | 18 | AAW29477  | Human histone H1 i  | 184 | 7 | 0.9 | 337 | 22 | ABW71412 | Drosophila melanog  |
| 112 | 7 | 0.9 | 218 | 20 | AAW34034  | Human histone H1 i  | 185 | 7 | 0.9 | 337 | 22 | ABW71412 | Kitsatosporia sp.   |
| 113 | 7 | 0.9 | 218 | 21 | AAW57332  | Human histone H1 i  | 186 | 7 | 0.9 | 340 | 20 | AAW14857 | 85-C. Mycobacteri   |
| 114 | 7 | 0.9 | 221 | 24 | ABU25464  | Aspergillus fumiga  | 187 | 7 | 0.9 | 340 | 20 | AAW14857 | Antigen 85C protei  |
| 115 | 7 | 0.9 | 225 | 21 | AAW01956  | Human secreted pro  | 188 | 7 | 0.9 | 340 | 20 | AAW47557 | Ag85C. Mycobacter   |
| 116 | 7 | 0.9 | 229 | 22 | AAW90579  | Human secreted pro  | 189 | 7 | 0.9 | 340 | 23 | ABW73463 | M tuberculosis 85C  |
| 117 | 7 | 0.9 | 229 | 23 | ABW65458  | Human albumin fusi  | 190 | 7 | 0.9 | 340 | 23 | AAW50731 | Mycobacterium tube  |
| 118 | 7 | 0.9 | 234 | 23 | AAW09944  | Human albumin fusi  | 191 | 7 | 0.9 | 345 | 12 | AAW14408 | Nuclear factor C/E  |
| 119 | 7 | 0.9 | 234 | 23 | AAW13234  | Histone H1-4, frag  | 192 | 7 | 0.9 | 345 | 12 | AAW14408 | Human CCAAT enhanc  |
| 120 | 7 | 0.9 | 245 | 20 | AAW42302  | Human linker histop | 193 | 7 | 0.9 | 345 | 22 | AAE11944 | Human CCAAT/enhanc  |
| 121 | 7 | 0.9 | 245 | 21 | AAW73441  | Trehalose-6-phosph  | 194 | 7 | 0.9 | 345 | 22 | AAE11944 | Human pancreatic c  |
| 122 | 7 | 0.9 | 245 | 21 | AAW67834  | Amino acid sequenc  | 195 | 7 | 0.9 | 345 | 23 | ABP68629 | Human pancreat      |
| 123 | 7 | 0.9 | 245 | 22 | ABW0547   | Human secreted pro  | 196 | 7 | 0.9 | 345 | 23 | AAO15441 | Human C/EBP-beta p  |
| 124 | 7 | 0.9 | 245 | 22 | ABW0577   | Human secreted pro  | 197 | 7 | 0.9 | 345 | 23 | AAO15442 | Human C/EBP-beta p  |
| 125 | 7 | 0.9 | 245 | 22 | AAW90578  | Human secreted pro  | 198 | 7 | 0.9 | 345 | 23 | AAO15443 | Human C/EBP-beta p  |
| 126 | 7 | 0.9 | 245 | 22 | AAW90580  | Human secreted pro  | 199 | 7 | 0.9 | 345 | 23 | AAO15449 | Human C/EBP-beta p  |
| 127 | 7 | 0.9 | 245 | 22 | AAW90581  | Human secreted pro  | 200 | 7 | 0.9 | 345 | 23 | AAO15450 | Human C/EBP-beta p  |
| 128 | 7 | 0.9 | 245 | 22 | AAW87413  | Human gene 5i enco  | 201 | 7 | 0.9 | 345 | 23 | AAO15451 | Human C/EBP-beta p  |
| 129 | 7 | 0.9 | 245 | 23 | ABG65399  | Human albumin fusi  | 202 | 7 | 0.9 | 345 | 23 | AAO15452 | Human C/EBP-beta p  |
| 130 | 7 | 0.9 | 245 | 23 | ABG65452  | Human albumin fusi  | 203 | 7 | 0.9 | 345 | 23 | AAO15453 | Human C/EBP-beta p  |
| 131 | 7 | 0.9 | 245 | 23 | ABG65454  | Human albumin fusi  | 204 | 7 | 0.9 | 345 | 23 | AAO15454 | Human C/EBP-beta m  |
| 132 | 7 | 0.9 | 245 | 23 | ABG65455  | Human albumin fusi  | 205 | 7 | 0.9 | 345 | 23 | AAU97259 | Alternative human   |
| 133 | 7 | 0.9 | 245 | 23 | ABG65456  | Human albumin fusi  | 206 | 7 | 0.9 | 345 | 23 | AAU97259 | Human alternative   |
| 134 | 7 | 0.9 | 245 | 23 | ABG65457  | Human albumin fusi  | 207 | 7 | 0.9 | 345 | 23 | AAU97261 | Human alternative   |
| 135 | 7 | 0.9 | 245 | 24 | ABW47660  | Human secreted pro  | 208 | 7 | 0.9 | 345 | 23 | AAU97262 | Wild-type human C/  |
| 136 | 7 | 0.9 | 245 | 24 | ABW47981  | Human secreted pro  | 209 | 7 | 0.9 | 345 | 23 | AAU97263 | Human C/EBP-beta pr |
| 137 | 7 | 0.9 | 250 | 23 | ABW54032  | Lactococcus lactis  | 210 | 7 | 0.9 | 345 | 23 | AAU97264 | Human C/EBP-beta pr |
| 138 | 7 | 0.9 | 267 | 22 | ABW404300 | Murine T-cell rece  | 211 | 7 | 0.9 | 345 | 23 | AAU97265 | Wild-type human C/  |
| 139 | 7 | 0.9 | 267 | 22 | ABW464636 | Drosophila melanog  | 212 | 7 | 0.9 | 345 | 23 | AAU97266 | Human C/EBP-beta pr |
| 140 | 7 | 0.9 | 270 | 22 | AGW33107  | C glutamicum prote  | 213 | 7 | 0.9 | 345 | 23 | AAU97267 | Human C/EBP-beta pr |
| 141 | 7 | 0.9 | 270 | 23 | ABU50901  | Helicobacter pylor  | 214 | 7 | 0.9 | 345 | 23 | AAU97267 | Human C/EBP-beta pr |
| 142 | 7 | 0.9 | 270 | 23 | ABW95678  | Human GPCR polypep  | 215 | 7 | 0.9 | 348 | 20 | AAW49182 | Human C/EBP-beta.   |
| 143 | 7 | 0.9 | 273 | 13 | AAW27505  | 5' rat brain somat  | 216 | 7 | 0.9 | 348 | 20 | AAW49182 | Amino acid sequenc  |
| 144 | 7 | 0.9 | 276 | 21 | AAW50947  | Arabidopsis thalia  | 217 | 7 | 0.9 | 352 | 20 | AAW98042 | M vaccae GV-24B pr  |
| 145 | 7 | 0.9 | 276 | 23 | ABW48710  | Listeria monocytog  | 218 | 7 | 0.9 | 352 | 19 | AAW98042 | Lactococcus lactis  |
| 146 | 7 | 0.9 | 277 | 21 | AAW05115  | Arabidopsis thalia  | 219 | 7 | 0.9 | 359 | 21 | AAW98649 | H. pylori GHP0 403  |
| 147 | 7 | 0.9 | 278 | 24 | AAW79677  | Arabidopsis thalia  | 220 | 7 | 0.9 | 359 | 23 | ABU52057 | Helicobacter pylor  |
| 148 | 7 | 0.9 | 278 | 24 | ABP26995  | Human ENZM-7, incy  | 221 | 7 | 0.9 | 359 | 23 | ABU52057 | Helicobacter pylor  |
| 149 | 7 | 0.9 | 279 | 23 | ABP26995  | Streptococcus poly  | 222 | 7 | 0.9 | 360 | 21 | AAW05582 | Arabidopsis thalia  |
| 150 | 7 | 0.9 | 285 | 21 | AAW21607  | Arabidopsis thalia  | 223 | 7 | 0.9 | 360 | 21 | AAW05582 | Arabidopsis thalia  |
| 151 | 7 | 0.9 | 285 | 22 | AAU305588 | Novel human secret  | 224 | 7 | 0.9 | 362 | 21 | AAW50872 | Arabidopsis thalia  |
| 152 | 7 | 0.9 | 287 | 21 | AAW113325 | Arabidopsis thalia  | 225 | 7 | 0.9 | 365 | 19 | AAW50943 | Human papillomavir  |
| 153 | 7 | 0.9 | 287 | 24 | ABJ26064  | Aspergillus fumiga  | 226 | 7 | 0.9 | 365 | 22 | AAW98425 | Human papillomavir  |
| 154 | 7 | 0.9 | 292 | 22 | AAW23629  | Human protein sequ  | 227 | 7 | 0.9 | 365 | 23 | ABW70652 | Human papillomavir  |
| 155 | 7 | 0.9 | 294 | 22 | ABW59210  | Drosophila melanog  | 228 | 7 | 0.9 | 368 | 19 | AAW68494 | E2 papillomavirus   |



|     |   |     |     |    |          |                     |     |   |     |      |    |          |                     |
|-----|---|-----|-----|----|----------|---------------------|-----|---|-----|------|----|----------|---------------------|
| 229 | 7 | 0.9 | 368 | 22 | AAB98439 | Human papillomavir  | 302 | 7 | 0.9 | 537  | 22 | ABB59734 | Drosophila melanog  |
| 230 | 7 | 0.9 | 370 | 21 | AAB56905 | Human prostate can  | 303 | 7 | 0.9 | 539  | 22 | AAU34904 | Enterococcus faeca  |
| 231 | 7 | 0.9 | 372 | 22 | ABG16669 | Novel human diagno  | 304 | 7 | 0.9 | 544  | 20 | AAU34762 | Chlamydia pneumoni  |
| 232 | 7 | 0.9 | 377 | 10 | AAU95118 | Fusion protein of   | 305 | 7 | 0.9 | 551  | 21 | AAU52398 | Human keratin KER1  |
| 233 | 7 | 0.9 | 377 | 10 | AAU93412 | Fusion of transfor  | 306 | 7 | 0.9 | 551  | 23 | AAE20423 | Human keratin-2 (K  |
| 234 | 7 | 0.9 | 377 | 21 | AAG55581 | Arabidopsis thalia  | 307 | 7 | 0.9 | 557  | 22 | ABE71470 | Drosophila melanog  |
| 235 | 7 | 0.9 | 377 | 21 | AAG50871 | Arabidopsis thalia  | 308 | 7 | 0.9 | 558  | 24 | ABP80490 | N. gonorrhoeae ami  |
| 236 | 7 | 0.9 | 377 | 21 | AAG50942 | Arabidopsis thalia  | 309 | 7 | 0.9 | 561  | 22 | AAU36080 | Klebsiella pneumoni |
| 237 | 7 | 0.9 | 378 | 21 | AAG19974 | Arabidopsis thalia  | 310 | 7 | 0.9 | 562  | 22 | AAU36524 | Human polyepitide   |
| 238 | 7 | 0.9 | 378 | 21 | AAG45204 | Arabidopsis thalia  | 311 | 7 | 0.9 | 568  | 22 | AAU39524 | Propionibacterium   |
| 239 | 7 | 0.9 | 380 | 21 | AAU42509 | Human ORFX ORP273   | 312 | 7 | 0.9 | 572  | 19 | AAU68486 | Mouse Ulf1-1 prote  |
| 240 | 7 | 0.9 | 388 | 20 | AAU04997 | Mycobacterium spec  | 313 | 7 | 0.9 | 572  | 23 | ABG32228 | Human Ulp2/CRMP2    |
| 241 | 7 | 0.9 | 388 | 23 | ABG60300 | Lymphona associat   | 314 | 7 | 0.9 | 578  | 23 | ABP73956 | Candida albicans e  |
| 242 | 7 | 0.9 | 388 | 23 | ABG03271 | G protein-coupled   | 315 | 7 | 0.9 | 589  | 24 | ABU00117 | Human novel polype  |
| 243 | 7 | 0.9 | 388 | 24 | ABP76558 | Amino acid sequenc  | 316 | 7 | 0.9 | 603  | 22 | ABU71538 | Drosophila melanog  |
| 244 | 7 | 0.9 | 390 | 22 | AAU99051 | Human somatostatin  | 317 | 7 | 0.9 | 603  | 22 | AAU38985 | Drosophila G-prote  |
| 245 | 7 | 0.9 | 391 | 14 | AAU33259 | Human somatostatin  | 318 | 7 | 0.9 | 610  | 22 | ABG55215 | Drosophila melanog  |
| 246 | 7 | 0.9 | 391 | 14 | AAU33260 | Murine somatostati  | 319 | 7 | 0.9 | 614  | 20 | AAU98112 | T-cell membrane pr  |
| 247 | 7 | 0.9 | 391 | 14 | AAU33260 | Non-endogenous hum  | 320 | 7 | 0.9 | 614  | 20 | AAU98113 | T-cell membrane pr  |
| 248 | 7 | 0.9 | 391 | 24 | ABP81913 | Human somatostatin  | 321 | 7 | 0.9 | 622  | 22 | ABG14288 | Novel human diagno  |
| 249 | 7 | 0.9 | 393 | 19 | AAU60133 | M. vaccae potd hom  | 322 | 7 | 0.9 | 622  | 20 | AAU98113 | Human novel cytol   |
| 250 | 7 | 0.9 | 393 | 20 | AAU14880 | M. vaccae potd gen  | 323 | 7 | 0.9 | 627  | 20 | AAU01495 | High level promote  |
| 251 | 7 | 0.9 | 393 | 23 | ABE73486 | C glutamicum prote  | 324 | 7 | 0.9 | 627  | 23 | ABG77914 | Rat pan-s/tk recep  |
| 252 | 7 | 0.9 | 394 | 22 | AAU90342 | C glutamicum prote  | 325 | 7 | 0.9 | 628  | 22 | AAU95279 | Human protein sequ  |
| 253 | 7 | 0.9 | 403 | 22 | AAU99734 | Human Doc2-beta co  | 326 | 7 | 0.9 | 630  | 22 | AAU34443 | Novel mar regulate  |
| 254 | 7 | 0.9 | 412 | 18 | AAU25032 | Human Doc2-beta co  | 327 | 7 | 0.9 | 630  | 22 | AAU29331 | Human protein sequ  |
| 255 | 7 | 0.9 | 417 | 22 | AAU04921 | Propionibacterium   | 328 | 7 | 0.9 | 632  | 22 | AAU94766 | Artichoke sucrose   |
| 256 | 7 | 0.9 | 419 | 22 | ABG22902 | Novel human diagno  | 329 | 7 | 0.9 | 637  | 19 | AAU77034 | Human protein sequ  |
| 257 | 7 | 0.9 | 420 | 21 | AAU45419 | Arabidopsis thalia  | 330 | 7 | 0.9 | 639  | 22 | AAU95818 | Amino acid sequenc  |
| 258 | 7 | 0.9 | 421 | 22 | ABE67110 | Drosophila melanog  | 331 | 7 | 0.9 | 643  | 20 | AAU14964 | C heterostrophus F  |
| 259 | 7 | 0.9 | 421 | 23 | AAU49674 | T. thermophilus HB  | 332 | 7 | 0.9 | 643  | 23 | ABG41524 | Human ovarian anti  |
| 260 | 7 | 0.9 | 426 | 16 | AAU74619 | Xanthomonas L-glut  | 333 | 7 | 0.9 | 653  | 23 | ABG06991 | Rat very long chai  |
| 261 | 7 | 0.9 | 432 | 19 | AAU98299 | H. pylori GHPO 109  | 334 | 7 | 0.9 | 653  | 23 | ABG06992 | Mouse very long ch  |
| 262 | 7 | 0.9 | 432 | 23 | ABG98549 | Human nucleic acid  | 335 | 7 | 0.9 | 655  | 23 | ABG06992 | Drosophila melanog  |
| 263 | 7 | 0.9 | 433 | 21 | AAU45418 | Arabidopsis thalia  | 336 | 7 | 0.9 | 663  | 22 | ABG59368 | Drosophila melanog  |
| 264 | 7 | 0.9 | 433 | 23 | ABG97421 | S. macromyceticus   | 337 | 7 | 0.9 | 668  | 16 | AAU81315 | Yeast TIR2. Sacch   |
| 265 | 7 | 0.9 | 433 | 24 | ABU58233 | Rice stress respon  | 338 | 7 | 0.9 | 668  | 19 | AAU54158 | S. cerevisiae TIR2  |
| 266 | 7 | 0.9 | 441 | 9  | AAU82998 | Tobacco GAPB incl   | 339 | 7 | 0.9 | 668  | 22 | AAU74488 | S. cerevisiae TIR2  |
| 267 | 7 | 0.9 | 442 | 21 | AAU45417 | Arabidopsis thalia  | 340 | 7 | 0.9 | 672  | 7  | AAU60219 | Sequence of thermo  |
| 268 | 7 | 0.9 | 445 | 22 | ABG60464 | Drosophila melanog  | 341 | 7 | 0.9 | 672  | 8  | AAU71200 | Thermostable beta-  |
| 269 | 7 | 0.9 | 445 | 22 | AAU41310 | Human polypeptide   | 342 | 7 | 0.9 | 682  | 22 | AAU63224 | Amino acid sequenc  |
| 270 | 7 | 0.9 | 446 | 22 | ABU11726 | Novel human diagno  | 343 | 7 | 0.9 | 682  | 22 | ABG57992 | Drosophila melanog  |
| 271 | 7 | 0.9 | 446 | 24 | ABU58173 | Human semaphorin Y  | 344 | 7 | 0.9 | 706  | 22 | AAU94174 | Human protein sequ  |
| 272 | 7 | 0.9 | 451 | 22 | AAU68528 | Corn stress respon  | 345 | 7 | 0.9 | 741  | 17 | AAU99327 | Membrane anchor pr  |
| 273 | 7 | 0.9 | 452 | 24 | ABU78391 | Human novel cytol   | 346 | 7 | 0.9 | 749  | 23 | AAU78526 | Rat STAI-1alpha am  |
| 274 | 7 | 0.9 | 452 | 24 | AAU78391 | N. gonorrhoeae ami  | 347 | 7 | 0.9 | 757  | 22 | AAU04600 | Oat Beta-amyrin sy  |
| 275 | 7 | 0.9 | 478 | 20 | AAU99599 | Human methionine a  | 348 | 7 | 0.9 | 757  | 22 | ABG15173 | Novel human diagno  |
| 276 | 7 | 0.9 | 478 | 20 | AAU93215 | Human p67 homolog   | 349 | 7 | 0.9 | 808  | 23 | ABG90957 | Herbicidally activ  |
| 277 | 7 | 0.9 | 478 | 20 | AAU94763 | Mouse type 2 methi  | 350 | 7 | 0.9 | 821  | 22 | ABG60968 | Drosophila melanog  |
| 278 | 7 | 0.9 | 478 | 20 | AAU94765 | Human type 2 methi  | 351 | 7 | 0.9 | 822  | 19 | AAU41943 | Herbicidally activ  |
| 279 | 7 | 0.9 | 478 | 22 | ABG50275 | elf-2-associated p  | 352 | 7 | 0.9 | 822  | 23 | ABG18111 | Human OC-116KDa     |
| 280 | 7 | 0.9 | 478 | 22 | ABG28377 | Human methionine a  | 353 | 7 | 0.9 | 833  | 21 | AAU42353 | Human ORFX ORP2117  |
| 281 | 7 | 0.9 | 478 | 24 | ABG76374 | Human type 2 amino  | 354 | 7 | 0.9 | 850  | 23 | ABG93031 | Herbicidally activ  |
| 282 | 7 | 0.9 | 478 | 24 | ABG76375 | Mouse type 2 amino  | 355 | 7 | 0.9 | 904  | 9  | AAU80914 | Sequence of Herpes  |
| 283 | 7 | 0.9 | 478 | 24 | ABG76377 | Human type 2 amino  | 356 | 7 | 0.9 | 904  | 14 | AAU41778 | Glycoprotein B (gB  |
| 284 | 7 | 0.9 | 478 | 24 | ABG76378 | Mouse type 2 amino  | 357 | 7 | 0.9 | 904  | 17 | AAU00376 | HSV-2 glycoprotein  |
| 285 | 7 | 0.9 | 478 | 24 | ABU07490 | Protein differenti  | 358 | 7 | 0.9 | 904  | 18 | AAU34552 | Herpes simplex vir  |
| 286 | 7 | 0.9 | 484 | 23 | AAU20303 | Chlamydia pneumoni  | 359 | 7 | 0.9 | 904  | 19 | AAU72193 | Herpes simplex vir  |
| 287 | 7 | 0.9 | 486 | 20 | AAU34696 | Chlamydia pneumoni  | 360 | 7 | 0.9 | 904  | 22 | AAU74442 | HSV-2 strain S95 C  |
| 288 | 7 | 0.9 | 486 | 23 | AAU20292 | Chlamydia pneumoni  | 361 | 7 | 0.9 | 904  | 23 | AAU17812 | Herpes simplex vir  |
| 289 | 7 | 0.9 | 493 | 22 | ABU70327 | Drosophila melanog  | 362 | 7 | 0.9 | 904  | 23 | AAU17812 | Human semaphorin Y  |
| 290 | 7 | 0.9 | 494 | 22 | ABG14289 | Novel human diagno  | 363 | 7 | 0.9 | 930  | 24 | AAU57260 | Human semaphorin Y  |
| 291 | 7 | 0.9 | 494 | 23 | AAU64944 | Human protein SRQ   | 364 | 7 | 0.9 | 934  | 24 | ABU72510 | Aspergillus fumiga  |
| 292 | 7 | 0.9 | 496 | 20 | AAU18653 | Chlamydia trachoma  | 365 | 7 | 0.9 | 959  | 24 | ABU72510 | Aspergillus fumiga  |
| 293 | 7 | 0.9 | 497 | 23 | AAU20296 | Chlamydia trachoma  | 366 | 7 | 0.9 | 998  | 22 | AAU96484 | Putative P. abyss   |
| 294 | 7 | 0.9 | 500 | 22 | AAU73991 | Human colon cancer  | 367 | 7 | 0.9 | 1005 | 23 | AAU59636 | Drosophila melanog  |
| 295 | 7 | 0.9 | 510 | 23 | AAU10966 | Xylella fastidiosa  | 368 | 7 | 0.9 | 1005 | 23 | AAU66143 | Drosophila PAS Kin  |
| 296 | 7 | 0.9 | 512 | 22 | ABG61369 | Drosophila melanog  | 369 | 7 | 0.9 | 1041 | 21 | AAU38594 | Arabidopsis thalia  |
| 297 | 7 | 0.9 | 514 | 21 | AAU37324 | Arabidopsis thalia  | 370 | 7 | 0.9 | 1041 | 21 | AAU11737 | Cryptosporidium pa  |
| 298 | 7 | 0.9 | 514 | 22 | AAU92523 | Human protein sequ  | 371 | 7 | 0.9 | 1053 | 24 | ABU48168 | Human bladder can   |
| 299 | 7 | 0.9 | 516 | 20 | AAU55528 | Human STLK3 protei  | 372 | 7 | 0.9 | 1095 | 21 | AAU01847 | Haemophilus influe  |
| 300 | 7 | 0.9 | 526 | 12 | AAU13247 | A. niger pyruvate k | 373 | 7 | 0.9 | 1118 | 21 | AAU18809 | A biogenic amine r  |
| 301 | 7 | 0.9 | 534 | 22 | AAU47622 | Synthetic lipi. C   | 374 | 7 | 0.9 | 1166 | 22 | ABU70258 | Drosophila melanog  |

|     |   |     |      |    |          |                     |     |   |     |    |    |          |                    |
|-----|---|-----|------|----|----------|---------------------|-----|---|-----|----|----|----------|--------------------|
| 375 | 7 | 0.9 | 1217 | 22 | ABB64764 | Drosophila melanog  | 448 | 6 | 0.8 | 9  | 23 | ABJ16146 | Zinc transporter p |
| 376 | 7 | 0.9 | 1229 | 22 | ABB65036 | Drosophila melanog  | 449 | 6 | 0.8 | 9  | 23 | ABJ16147 | Zinc transporter p |
| 377 | 7 | 0.9 | 1279 | 23 | ABG70787 | Human kinesin-rela  | 450 | 6 | 0.8 | 9  | 23 | ABJ16162 | Zinc transporter p |
| 378 | 7 | 0.9 | 1279 | 23 | ABB80078 | Human kinesin moto  | 451 | 6 | 0.8 | 9  | 23 | ABJ16164 | Zinc transporter p |
| 379 | 7 | 0.9 | 1279 | 24 | ABG72397 | Human partial kine  | 452 | 6 | 0.8 | 9  | 23 | ABJ16165 | Zinc transporter p |
| 380 | 7 | 0.9 | 1294 | 22 | ABB63502 | Drosophila melanog  | 453 | 6 | 0.8 | 9  | 23 | ABJ16313 | Zinc transporter p |
| 381 | 7 | 0.9 | 1314 | 22 | ABG22645 | Novel human diagn   | 454 | 6 | 0.8 | 9  | 23 | ABJ16423 | Zinc transporter p |
| 382 | 7 | 0.9 | 1373 | 23 | ABP73292 | Candida albicans e  | 455 | 6 | 0.8 | 9  | 23 | ABJ16752 | Zinc transporter p |
| 383 | 7 | 0.9 | 1406 | 22 | AAU32728 | Novel human secret  | 456 | 6 | 0.8 | 9  | 23 | ABJ17445 | Zinc transporter p |
| 384 | 7 | 0.9 | 1514 | 22 | ABP70827 | Drosophila melanog  | 457 | 6 | 0.8 | 9  | 23 | ABJ17448 | Human cancer-relat |
| 385 | 7 | 0.9 | 1536 | 14 | AA41723  | High molecular wei  | 458 | 6 | 0.8 | 9  | 24 | ABR12005 | Human cancer-relat |
| 386 | 7 | 0.9 | 1536 | 14 | AA41725  | High molecular wei  | 459 | 6 | 0.8 | 9  | 24 | ABR12814 | Human cancer-relat |
| 387 | 7 | 0.9 | 1536 | 15 | AA63505  | Haemophilus high m  | 460 | 6 | 0.8 | 9  | 24 | ABR13032 | HRV2 2A proteinase |
| 388 | 7 | 0.9 | 1536 | 18 | AAW30293 | Non-typesable Haemo | 461 | 6 | 0.8 | 10 | 14 | ABR13038 | HIV A01 super moti |
| 389 | 7 | 0.9 | 1536 | 21 | AAW30293 | Haemophilus influe  | 462 | 6 | 0.8 | 10 | 22 | ABR13038 | HIV A02 super moti |
| 390 | 7 | 0.9 | 1536 | 21 | AAW30293 | Haemophilus influe  | 463 | 6 | 0.8 | 10 | 22 | ABR13038 | HIV A24 super moti |
| 391 | 7 | 0.9 | 1721 | 19 | AAW48299 | Cryptosporidium pa  | 464 | 6 | 0.8 | 10 | 22 | ABR13038 | HIV A03 motif env  |
| 392 | 7 | 0.9 | 1721 | 21 | AAW48299 | Portion of Cryptos  | 465 | 6 | 0.8 | 10 | 22 | ABR13038 | HIV A03 motif env  |
| 393 | 7 | 0.9 | 1721 | 21 | AAW48299 | C parvum GP900 pro  | 466 | 6 | 0.8 | 10 | 22 | ABR13038 | HIV A03 motif env  |
| 394 | 7 | 0.9 | 1861 | 21 | AAW90350 | Drosophila melanog  | 467 | 6 | 0.8 | 10 | 22 | ABR13038 | HIV A03 motif env  |
| 395 | 7 | 0.9 | 1861 | 22 | ABG71293 | Drosophila melanog  | 468 | 6 | 0.8 | 10 | 22 | ABR13038 | HIV A03 motif env  |
| 396 | 7 | 0.9 | 1861 | 22 | ABG71293 | Drosophila melanog  | 469 | 6 | 0.8 | 10 | 22 | ABR13038 | HIV A03 motif env  |
| 397 | 7 | 0.9 | 2069 | 22 | ABG70867 | C glutamicum prote  | 470 | 6 | 0.8 | 10 | 22 | ABR13038 | Human complementar |
| 398 | 7 | 0.9 | 2273 | 22 | ABG70867 | Novel human diagn   | 471 | 6 | 0.8 | 10 | 22 | ABR13038 | Arabidopsis thalia |
| 399 | 7 | 0.9 | 2703 | 23 | ABG63299 | Drosophila melanog  | 472 | 6 | 0.8 | 10 | 22 | ABR13038 | Arabidopsis thalia |
| 400 | 7 | 0.9 | 2703 | 23 | ABG70019 | Larval viability a  | 473 | 6 | 0.8 | 10 | 22 | ABR13038 | Arabidopsis thalia |
| 401 | 7 | 0.9 | 3129 | 21 | AAW07427 | Amino acid sequenc  | 474 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 402 | 7 | 0.9 | 3170 | 20 | AAV39299 | SpnC a polyketide   | 475 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 403 | 7 | 0.9 | 3536 | 22 | ABE65480 | S. spinosa protein  | 476 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 404 | 7 | 0.9 | 3562 | 22 | ABE65480 | Drosophila melanog  | 477 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 405 | 7 | 0.9 | 4654 | 22 | ABG19817 | Polyketide synthas  | 478 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 406 | 7 | 0.9 | 4718 | 22 | ABG11811 | Novel human diagn   | 479 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 407 | 6 | 0.8 | 6    | 20 | AAW87909 | Novel human diagn   | 480 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 408 | 6 | 0.8 | 6    | 20 | AAW87909 | Novel human diagn   | 481 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 409 | 6 | 0.8 | 7    | 21 | AAV62993 | Leader peptide use  | 482 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 410 | 6 | 0.8 | 7    | 23 | AAE29596 | Metallopeptide N-t  | 483 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 411 | 6 | 0.8 | 7    | 23 | AAE29596 | PB-cadherin cell a  | 484 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 412 | 6 | 0.8 | 8    | 14 | AAK38553 | Metallopeptide N-t  | 485 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 413 | 6 | 0.8 | 8    | 21 | AAV62874 | Peptide #3 for tre  | 486 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 414 | 6 | 0.8 | 8    | 21 | AAV62874 | PB-cadherin cell a  | 487 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 415 | 6 | 0.8 | 8    | 21 | AAV62874 | PB-cadherin cell a  | 488 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 416 | 6 | 0.8 | 8    | 22 | ABP14271 | PB-cadherin cell a  | 489 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 417 | 6 | 0.8 | 8    | 22 | ABP14271 | HIV A03 super moti  | 490 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 418 | 6 | 0.8 | 8    | 22 | ABP19801 | HIV A03 motif env   | 491 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 419 | 6 | 0.8 | 8    | 22 | ABP22280 | HIV A03 motif env   | 492 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 420 | 6 | 0.8 | 8    | 23 | AAE29586 | Metallopeptide N-t  | 493 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 421 | 6 | 0.8 | 9    | 21 | AAV62875 | Metallopeptide N-t  | 494 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 422 | 6 | 0.8 | 9    | 21 | AAV62875 | PB-cadherin cell a  | 495 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 423 | 6 | 0.8 | 9    | 21 | AAV62875 | PB-cadherin cell a  | 496 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 424 | 6 | 0.8 | 9    | 22 | ABP12149 | PB-cadherin cell a  | 497 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 425 | 6 | 0.8 | 9    | 22 | ABP12149 | HIV A02 super moti  | 498 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 426 | 6 | 0.8 | 9    | 22 | ABP19808 | HIV A03 motif env   | 499 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 427 | 6 | 0.8 | 9    | 22 | AAW23250 | HIV peptide SEQ ID  | 500 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 428 | 6 | 0.8 | 9    | 23 | AAE29579 | HIV peptide SEQ ID  | 501 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 429 | 6 | 0.8 | 9    | 23 | AAE29579 | Metallopeptide N-t  | 502 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 430 | 6 | 0.8 | 9    | 23 | AAE29585 | Metallopeptide N-t  | 503 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 431 | 6 | 0.8 | 9    | 23 | ABJ15347 | Zinc transporter p  | 504 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 432 | 6 | 0.8 | 9    | 23 | ABJ15754 | Zinc transporter p  | 505 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 433 | 6 | 0.8 | 9    | 23 | ABJ16001 | Zinc transporter p  | 506 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 434 | 6 | 0.8 | 9    | 23 | ABJ16053 | Zinc transporter p  | 507 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 435 | 6 | 0.8 | 9    | 23 | ABJ16067 | Zinc transporter p  | 508 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 436 | 6 | 0.8 | 9    | 23 | ABJ16069 | Zinc transporter p  | 509 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 437 | 6 | 0.8 | 9    | 23 | ABJ16070 | Zinc transporter p  | 510 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 438 | 6 | 0.8 | 9    | 23 | ABJ16086 | Zinc transporter p  | 511 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 439 | 6 | 0.8 | 9    | 23 | ABJ16087 | Zinc transporter p  | 512 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 440 | 6 | 0.8 | 9    | 23 | ABJ16091 | Zinc transporter p  | 513 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 441 | 6 | 0.8 | 9    | 23 | ABJ16105 | Zinc transporter p  | 514 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 442 | 6 | 0.8 | 9    | 23 | ABJ16107 | Zinc transporter p  | 515 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 443 | 6 | 0.8 | 9    | 23 | ABJ16111 | Zinc transporter p  | 516 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 444 | 6 | 0.8 | 9    | 23 | ABJ16123 | Zinc transporter p  | 517 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 445 | 6 | 0.8 | 9    | 23 | ABJ16137 | Zinc transporter p  | 518 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 446 | 6 | 0.8 | 9    | 23 | ABJ16138 | Zinc transporter p  | 519 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |
| 447 | 6 | 0.8 | 9    | 23 | ABJ16143 | Zinc transporter p  | 520 | 6 | 0.8 | 10 | 22 | ABR13038 | Saccharomyces cere |

|     |   |     |    |    |           |                     |     |    |    |           |                     |
|-----|---|-----|----|----|-----------|---------------------|-----|----|----|-----------|---------------------|
| 521 | 6 | 0.8 | 10 | 23 | ABJ17451  | Zinc transporter p  | 594 | 15 | 23 | ABJ17551  | Zinc transporter p  |
| 522 | 6 | 0.8 | 10 | 23 | ABJ17453  | Zinc transporter p  | 595 | 15 | 23 | ABJ17575  | Zinc transporter p  |
| 523 | 6 | 0.8 | 10 | 23 | ABJ17454  | Zinc transporter p  | 596 | 15 | 23 | ABJ17583  | Zinc transporter p  |
| 524 | 6 | 0.8 | 10 | 24 | ABJ12095  | Human cancer-relat  | 597 | 15 | 23 | ABJ17595  | Zinc transporter p  |
| 525 | 6 | 0.8 | 10 | 24 | ABR12052  | Human cancer-relat  | 598 | 15 | 23 | ABJ17647  | Zinc transporter p  |
| 526 | 6 | 0.8 | 10 | 24 | ABR12559  | Human cancer-relat  | 599 | 15 | 23 | ABJ17713  | Zinc transporter p  |
| 527 | 6 | 0.8 | 10 | 24 | ABR12728  | Human cancer-relat  | 600 | 15 | 23 | ABJ17714  | Zinc transporter p  |
| 528 | 6 | 0.8 | 10 | 24 | ABR12887  | Human cancer-relat  | 601 | 15 | 23 | ABJ17755  | Zinc transporter p  |
| 529 | 6 | 0.8 | 10 | 24 | ABR12914  | Human cancer-relat  | 602 | 15 | 23 | ABJ17883  | Zinc transporter p  |
| 530 | 6 | 0.8 | 11 | 20 | AAV25091  | Transduction prote  | 603 | 15 | 23 | ABJ17915  | Zinc transporter p  |
| 531 | 6 | 0.8 | 11 | 20 | AAV20867  | Expression constru  | 604 | 15 | 23 | ABJ17916  | Zinc transporter p  |
| 532 | 6 | 0.8 | 11 | 20 | AAV08780  | Expression constru  | 605 | 15 | 23 | ABJ17917  | Zinc transporter p  |
| 533 | 6 | 0.8 | 11 | 21 | AAAB29441 | Synthetic transduc  | 606 | 15 | 23 | ABJ17919  | Zinc transporter p  |
| 534 | 6 | 0.8 | 11 | 21 | AAV93557  | Amino acid sequenc  | 607 | 15 | 23 | ABJ17920  | Zinc transporter p  |
| 535 | 6 | 0.8 | 11 | 23 | ABP17604  | HIV B58 super moti  | 608 | 15 | 23 | ABJ17921  | Zinc transporter p  |
| 536 | 6 | 0.8 | 11 | 23 | AAE29570  | Metallopeptide N-t  | 609 | 15 | 23 | ABJ17923  | Zinc transporter p  |
| 537 | 6 | 0.8 | 11 | 24 | ABP56091  | Protein transducti  | 610 | 15 | 23 | ABJ17924  | Zinc transporter p  |
| 538 | 6 | 0.8 | 12 | 18 | AAW45220  | Vasoaactive intesti | 611 | 15 | 23 | ABJ17925  | Zinc transporter p  |
| 539 | 6 | 0.8 | 12 | 19 | AAW56722  | Synthetic peptide   | 612 | 15 | 23 | ABJ17927  | Zinc transporter p  |
| 540 | 6 | 0.8 | 12 | 19 | AAW56723  | Synthetic peptide   | 613 | 15 | 23 | ABJ17929  | Zinc transporter p  |
| 541 | 6 | 0.8 | 12 | 19 | AAW86724  | Synthetic peptide   | 614 | 15 | 24 | ABR33101  | Human cancer-relat  |
| 542 | 6 | 0.8 | 12 | 19 | AAW86712  | Synthetic peptide   | 615 | 15 | 24 | ABR33285  | Human cancer-relat  |
| 543 | 6 | 0.8 | 13 | 13 | AAAR22001 | Polypeptide A base  | 616 | 15 | 24 | ABR33291  | Human cancer-relat  |
| 544 | 6 | 0.8 | 13 | 14 | AAAR41800 | Thrombomodulin pep  | 617 | 15 | 24 | ABR33305  | Human cancer-relat  |
| 545 | 6 | 0.8 | 13 | 14 | AAAB15876 | Human chemokine de  | 618 | 15 | 24 | ABR33337  | Human cancer-relat  |
| 546 | 6 | 0.8 | 13 | 21 | AAV87758  | Potato THT protein  | 619 | 15 | 24 | ABR33375  | Human cancer-relat  |
| 547 | 6 | 0.8 | 13 | 21 | AAV87773  | Potato THT peptide  | 620 | 15 | 24 | ABR33406  | Human cancer-relat  |
| 548 | 6 | 0.8 | 13 | 22 | AAAG62659 | Prion beta-sheet b  | 621 | 15 | 24 | ABR33410  | Human cancer-relat  |
| 549 | 6 | 0.8 | 13 | 23 | AAE29557  | Metallopeptide #1   | 622 | 15 | 24 | ABP82663  | Human cancer-relat  |
| 550 | 6 | 0.8 | 13 | 23 | AAE27411  | Human granulocyte   | 623 | 15 | 24 | ABP74811  | G protease analysis |
| 551 | 6 | 0.8 | 13 | 23 | AAE27412  | Human granulocyte   | 624 | 16 | 14 | AAAR37202 | HRV2 2A proteinase  |
| 552 | 6 | 0.8 | 13 | 23 | AAE27413  | Human granulocyte   | 625 | 16 | 14 | AAAR43119 | Rhinoviral HRV2 2A  |
| 553 | 6 | 0.8 | 13 | 23 | AAE30607  | Human granulocyte   | 626 | 16 | 17 | AAAR89953 | eIF-4-gamma homolo  |
| 554 | 6 | 0.8 | 13 | 24 | AAE30607  | Human granulocyte   | 627 | 16 | 23 | AAU93316  | Human Rhinovirus (  |
| 555 | 6 | 0.8 | 13 | 24 | AAE30620  | Human granulocyte   | 628 | 16 | 24 | ABR48313  | Peptide SEQ ID NO:  |
| 556 | 6 | 0.8 | 13 | 24 | AAE30633  | Human granulocyte   | 629 | 16 | 24 | ABR48370  | Peptide SEQ ID NO:  |
| 557 | 6 | 0.8 | 13 | 24 | AAE30645  | Human granulocyte   | 630 | 17 | 22 | AAAB50045 | Human TREK peptide  |
| 558 | 6 | 0.8 | 14 | 18 | AAW19090  | Trypanosoma cruzi   | 631 | 17 | 23 | AAU90041  | Insulin/insulin-l   |
| 559 | 6 | 0.8 | 14 | 18 | AAW19091  | Trypanosoma cruzi   | 632 | 18 | 15 | AAAR47796 | HIV epitope #127    |
| 560 | 6 | 0.8 | 14 | 20 | AAV32846  | TCE repeat sequenc  | 633 | 18 | 20 | AAV05386  | Mouse GCR9 protein  |
| 561 | 6 | 0.8 | 14 | 20 | AAV32847  | TCE repeat sequenc  | 634 | 20 | 22 | AAAG63438 | Amino acid sequenc  |
| 562 | 6 | 0.8 | 14 | 20 | AAV23346  | Repeat sequence of  | 635 | 20 | 22 | AAAG63439 | Amino acid sequenc  |
| 563 | 6 | 0.8 | 14 | 20 | AAV23317  | Repeat sequence of  | 636 | 20 | 22 | AAAG63440 | Amino acid sequenc  |
| 564 | 6 | 0.8 | 14 | 21 | AAAB26471 | T.cruzi synthetic   | 637 | 20 | 22 | AAAG63441 | Amino acid sequenc  |
| 565 | 6 | 0.8 | 14 | 21 | AAAB26472 | T.cruzi synthetic   | 638 | 20 | 22 | AAAG63443 | Amino acid sequenc  |
| 566 | 6 | 0.8 | 14 | 21 | AAV92958  | Transforming growt  | 639 | 20 | 22 | AAAG63445 | Amino acid sequenc  |
| 567 | 6 | 0.8 | 14 | 21 | AAV92959  | Transforming growt  | 640 | 20 | 22 | AAAG63446 | Amino acid sequenc  |
| 568 | 6 | 0.8 | 14 | 21 | AAV87737  | Potato THT protein  | 641 | 20 | 22 | AAAG63448 | Amino acid sequenc  |
| 569 | 6 | 0.8 | 14 | 23 | AAE29558  | Metallopeptide #2   | 642 | 20 | 22 | AAAG63449 | Amino acid sequenc  |
| 570 | 6 | 0.8 | 14 | 23 | AAE29559  | Metallopeptide #3   | 643 | 20 | 22 | AAAG63450 | Amino acid sequenc  |
| 571 | 6 | 0.8 | 14 | 23 | AAE29560  | Metallopeptide #3   | 644 | 20 | 22 | AAAG63451 | Amino acid sequenc  |
| 572 | 6 | 0.8 | 14 | 23 | AAE29561  | Metallopeptide #5   | 645 | 20 | 22 | AAAG63453 | Amino acid sequenc  |
| 573 | 6 | 0.8 | 14 | 23 | AAU99562  | Metallopeptide #6   | 646 | 20 | 22 | AAAG63455 | Amino acid sequenc  |
| 574 | 6 | 0.8 | 14 | 23 | AAU99695  | Peptide encoded by  | 647 | 20 | 22 | AAAG63456 | Amino acid sequenc  |
| 575 | 6 | 0.8 | 14 | 24 | ABU06308  | Maize GBSS glycosy  | 648 | 20 | 22 | AAU01428  | Peptide #14 derive  |
| 576 | 6 | 0.8 | 14 | 24 | ABU06309  | Maize GBSS glycosy  | 649 | 20 | 22 | ABP30904  | OES antibody epit   |
| 577 | 6 | 0.8 | 14 | 24 | ABU06312  | Maize GBSS glycosy  | 650 | 20 | 23 | AAU88234  | Insulin/insulin-l   |
| 578 | 6 | 0.8 | 14 | 24 | ABU06317  | Maize GBSS C-termi  | 651 | 20 | 23 | AAU88403  | Insulin/insulin-l   |
| 579 | 6 | 0.8 | 15 | 12 | AAAR11735 | Human Ig light cha  | 652 | 21 | 17 | AAW06914  | T.cruzi epitope T   |
| 580 | 6 | 0.8 | 15 | 14 | AAAR44439 | Laminin A chain de  | 653 | 21 | 18 | AAW19062  | Trypanosoma cruzi   |
| 581 | 6 | 0.8 | 15 | 14 | AAAR43120 | Rhinoviral HRV2 2A  | 654 | 21 | 18 | AAW19063  | Trypanosoma cruzi   |
| 582 | 6 | 0.8 | 15 | 14 | AAAR43121 | Rhinoviral HRV2 2A  | 655 | 21 | 18 | AAW19087  | Trypanosoma cruzi   |
| 583 | 6 | 0.8 | 15 | 14 | AAAR43124 | Rhinoviral HRV2 2A  | 656 | 21 | 19 | AAW60095  | M. vaccae antigen   |
| 584 | 6 | 0.8 | 15 | 14 | AAAR43125 | Rhinoviral HRV2 2A  | 657 | 21 | 20 | AAV29696  | Trypanosoma cruzi   |
| 585 | 6 | 0.8 | 15 | 14 | AAAR43126 | Rhinoviral HRV2 2A  | 658 | 21 | 20 | AAV32849  | TCE repeat sequenc  |
| 586 | 6 | 0.8 | 15 | 14 | AAAR43127 | Rhinoviral HRV2 2A  | 659 | 21 | 20 | AAV32839  | TCE repeat sequenc  |
| 587 | 6 | 0.8 | 15 | 16 | AAAR80519 | Secretion protein   | 660 | 21 | 20 | AAV32843  | TCE repeat sequenc  |
| 588 | 6 | 0.8 | 15 | 20 | AAW92583  | Mouse beta-actin a  | 661 | 21 | 20 | AAV14833  | N-terminal fragmen  |
| 589 | 6 | 0.8 | 15 | 20 | AAW92584  | Mouse beta-actin a  | 662 | 21 | 20 | AAV14833  | N-terminal fragmen  |
| 590 | 6 | 0.8 | 15 | 20 | AAW87801  | N-terminal sequenc  | 663 | 21 | 20 | AAV23310  | Epitope of the TCE  |
| 591 | 6 | 0.8 | 15 | 21 | AAAB19236 | N-terminal sequenc  | 664 | 21 | 20 | AAV23311  | Epitope of the TCE  |
| 592 | 6 | 0.8 | 15 | 23 | ABJ17478  | Zinc transporter p  | 665 | 21 | 20 | AAAB26465 | Repeat sequence of  |
| 593 | 6 | 0.8 | 15 | 23 | ABJ17496  | Zinc transporter p  | 666 | 21 | 21 | AAAB26466 | T.cruzi TCE antige  |

|     |    |    |          |                    |     |   |     |    |          |                    |
|-----|----|----|----------|--------------------|-----|---|-----|----|----------|--------------------|
| 667 | 21 | 21 | AAB26468 | T. cruzi synthetic | 740 | 6 | 0.8 | 22 | AAM28982 | Peptide #3019 enco |
| 668 | 21 | 22 | ABU52903 | Human uterus-deriv | 741 | 6 | 0.8 | 22 | AAM04218 | Peptide #2900 enco |
| 669 | 21 | 22 | AAU93821 | T. cruzi B cell ep | 742 | 6 | 0.8 | 22 | ABN91176 | Pancreatic hormone |
| 670 | 21 | 22 | ABB73439 | M vaccae antigen G | 743 | 6 | 0.8 | 22 | ABG38256 | Human peptide enco |
| 671 | 21 | 24 | ABU54416 | Alanine/proline ri | 744 | 6 | 0.8 | 23 | ABP30321 | OES antibody epit  |
| 672 | 21 | 21 | ABU12982 | Cysteine containin | 745 | 6 | 0.8 | 23 | ABP30321 | B. aphidicola ribo |
| 673 | 21 | 22 | ABG63478 | Amino acid sequenc | 746 | 6 | 0.8 | 21 | AAI77867 | Peptide #10215 enc |
| 674 | 21 | 22 | ABG63479 | Amino acid sequenc | 747 | 6 | 0.8 | 22 | AAI77867 | Protein encoded by |
| 675 | 21 | 22 | ABG63480 | Amino acid sequenc | 748 | 6 | 0.8 | 22 | AAI77867 | Amino acid sequenc |
| 676 | 21 | 22 | ABG63481 | Amino acid sequenc | 749 | 6 | 0.8 | 22 | ABP30300 | HIV-1 isolate IIIB |
| 677 | 21 | 22 | ABG63483 | Amino acid sequenc | 750 | 6 | 0.8 | 22 | ABP30300 | Gene 19 human secr |
| 678 | 21 | 22 | ABG63485 | Amino acid sequenc | 751 | 6 | 0.8 | 22 | ABP30300 | HIV-1 isolate IIIB |
| 679 | 21 | 22 | ABG63486 | Amino acid sequenc | 752 | 6 | 0.8 | 22 | ABP30300 | Human secreted pro |
| 680 | 21 | 22 | ABG63488 | Amino acid sequenc | 753 | 6 | 0.8 | 22 | ABP30300 | Human secreted pro |
| 681 | 21 | 22 | ABG63490 | Amino acid sequenc | 754 | 6 | 0.8 | 22 | ABP30300 | Human secreted pro |
| 682 | 21 | 22 | ABG63491 | Amino acid sequenc | 755 | 6 | 0.8 | 22 | ABP30300 | Bacteriophage Dp-1 |
| 683 | 21 | 22 | ABG63492 | Amino acid sequenc | 756 | 6 | 0.8 | 22 | ABP30300 | Fusarium poae hydr |
| 684 | 21 | 22 | ABG63493 | Amino acid sequenc | 757 | 6 | 0.8 | 22 | ABP30300 | Arabidopsis thalia |
| 685 | 21 | 22 | ABG63495 | Amino acid sequenc | 758 | 6 | 0.8 | 22 | ABP30300 | Human PER1 peptide |
| 686 | 21 | 22 | ABG63496 | Amino acid sequenc | 759 | 6 | 0.8 | 22 | ABP30300 | Mouse homologue of |
| 687 | 21 | 22 | ABG63497 | Human colon cancer | 760 | 6 | 0.8 | 22 | ABP30300 | N. gonorrhoeae ami |
| 688 | 21 | 24 | ABE33405 | PD-1-related pepti | 761 | 6 | 0.8 | 22 | ABP30300 | N-terminal of Vari |
| 689 | 21 | 24 | ABU18533 | Human secreted pro | 762 | 6 | 0.8 | 22 | ABP30300 | Human liver peptid |
| 690 | 21 | 22 | ABG63490 | Eubacterial DnaE2  | 763 | 6 | 0.8 | 22 | ABP30300 | Peptide #8591 enco |
| 691 | 21 | 22 | ABG63491 | Human liver peptid | 764 | 6 | 0.8 | 22 | ABP30300 | Protein #7144 enco |
| 692 | 21 | 22 | ABG63492 | Cysteine-containin | 765 | 6 | 0.8 | 22 | ABP30300 | Human brain expres |
| 693 | 21 | 22 | ABG63493 | Eubacterial DnaE1  | 766 | 6 | 0.8 | 22 | ABP30300 | Human bone marrow  |
| 694 | 21 | 22 | ABG63494 | Eubacterial DnaE1  | 767 | 6 | 0.8 | 22 | ABP30300 | Peptide #8898 enco |
| 695 | 21 | 22 | ABG63495 | Eubacterial DnaE1  | 768 | 6 | 0.8 | 22 | ABP30300 | Human peptide enco |
| 696 | 21 | 22 | ABG63496 | Eubacterial DnaE1  | 769 | 6 | 0.8 | 22 | ABP30300 | Human polyepitide  |
| 697 | 21 | 22 | ABG63497 | Eubacterial DnaE1  | 770 | 6 | 0.8 | 22 | ABP30300 | Peptide #4695 enco |
| 698 | 21 | 22 | ABG63498 | Eubacterial DnaE1  | 771 | 6 | 0.8 | 22 | ABP30300 | Peptide #4557 enco |
| 699 | 21 | 22 | ABG63499 | Eubacterial DnaE1  | 772 | 6 | 0.8 | 22 | ABP30300 | Streptococcus poly |
| 700 | 21 | 22 | ABG63500 | Eubacterial DnaE1  | 773 | 6 | 0.8 | 22 | ABP30300 | N-terminus of IGE  |
| 701 | 21 | 22 | ABG63501 | Peptide from the 3 | 774 | 6 | 0.8 | 22 | ABP30300 | Human secretory po |
| 702 | 21 | 22 | ABG63502 | Amino acid sequenc | 775 | 6 | 0.8 | 22 | ABP30300 | Human liver peptid |
| 703 | 21 | 22 | ABG63503 | Peptide from the 3 | 776 | 6 | 0.8 | 22 | ABP30300 | Peptide #3038 enco |
| 704 | 21 | 22 | ABG63504 | Peptide from the 3 | 777 | 6 | 0.8 | 22 | ABP30300 | Peptide #3066 enco |
| 705 | 21 | 22 | ABG63505 | Amino acid sequenc | 778 | 6 | 0.8 | 22 | ABP30300 | Protein #2987 enco |
| 706 | 21 | 22 | ABG63506 | Amino acid sequenc | 779 | 6 | 0.8 | 22 | ABP30300 | Human brain expres |
| 707 | 21 | 22 | ABG63507 | Amino acid sequenc | 780 | 6 | 0.8 | 22 | ABP30300 | Human bone marrow  |
| 708 | 21 | 22 | ABG63508 | Finger 1 domain of | 781 | 6 | 0.8 | 22 | ABP30300 | Peptide #3012 enco |
| 709 | 21 | 22 | ABG63509 | Finger 1 domain of | 782 | 6 | 0.8 | 22 | ABP30300 | Peptide #3098 enco |
| 710 | 21 | 22 | ABG63510 | Human polypeptide  | 783 | 6 | 0.8 | 22 | ABP30300 | Peptide #2973 enco |
| 711 | 21 | 22 | ABG63511 | Bacillus thuringie | 784 | 6 | 0.8 | 22 | ABP30300 | Gene 23 human secr |
| 712 | 21 | 22 | ABG63512 | Partial N-terminal | 785 | 6 | 0.8 | 22 | ABP30300 | C. parvum p23 prot |
| 713 | 21 | 22 | ABG63513 | N-terminus of Cart | 786 | 6 | 0.8 | 22 | ABP30300 | Human secreted pro |
| 714 | 21 | 22 | ABG63514 | Human cartilage in | 787 | 6 | 0.8 | 22 | ABP30300 | Pre-transforming g |
| 715 | 21 | 22 | ABG63515 | Cartilage inducing | 788 | 6 | 0.8 | 22 | ABP30300 | Protonibacterium   |
| 716 | 21 | 22 | ABG63516 | Cartilage inducing | 789 | 6 | 0.8 | 22 | ABP30300 | Novel human design |
| 717 | 21 | 22 | ABG63517 | Cartilage-inducing | 790 | 6 | 0.8 | 22 | ABP30300 | Sequence of genom  |
| 718 | 21 | 22 | ABG63518 | Glucagon-like pept | 791 | 6 | 0.8 | 22 | ABP30300 | Human pre-TGF-beta |
| 719 | 21 | 22 | ABG63519 | Arg-specific gingi | 792 | 6 | 0.8 | 22 | ABP30300 | Human liver peptid |
| 720 | 21 | 22 | ABG63520 | Peptide from the 3 | 793 | 6 | 0.8 | 22 | ABP30300 | Human liver peptid |
| 721 | 21 | 22 | ABG63521 | Peptide from the 3 | 794 | 6 | 0.8 | 22 | ABP30300 | Protonibacterium   |
| 722 | 21 | 22 | ABG63522 | Peptide from the 3 | 795 | 6 | 0.8 | 22 | ABP30300 | Peptide #4943 enco |
| 723 | 21 | 22 | ABG63523 | Peptide from the 3 | 796 | 6 | 0.8 | 22 | ABP30300 | Peptide #11835 enc |
| 724 | 21 | 22 | ABG63524 | Human polypeptide  | 797 | 6 | 0.8 | 22 | ABP30300 | Human bone marrow  |
| 725 | 21 | 22 | ABG63525 | Partial peptide #4 | 798 | 6 | 0.8 | 22 | ABP30300 | Peptide #11836 enc |
| 726 | 21 | 22 | ABG63526 | N. gonorrhoeae ami | 799 | 6 | 0.8 | 22 | ABP30300 | Human peptide enco |
| 727 | 21 | 22 | ABG63527 | Peptide fragment o | 800 | 6 | 0.8 | 22 | ABP30300 | Human peptide enco |
| 728 | 21 | 22 | ABG63528 | V-3 loop region.   | 801 | 6 | 0.8 | 22 | ABP30300 | Arabidopsis thalia |
| 729 | 21 | 22 | ABG63529 | Human nervous syst | 802 | 6 | 0.8 | 22 | ABP30300 | Human liver peptid |
| 730 | 21 | 22 | ABG63530 | Human liver secret | 803 | 6 | 0.8 | 22 | ABP30300 | Protonibacterium   |
| 731 | 21 | 22 | ABG63531 | Human novel peptid | 804 | 6 | 0.8 | 22 | ABP30300 | Peptide #2077 enco |
| 732 | 21 | 22 | ABG63532 | Human liver peptid | 805 | 6 | 0.8 | 22 | ABP30300 | Protein #1975 enco |
| 733 | 21 | 22 | ABG63533 | Peptide #2960 enco | 806 | 6 | 0.8 | 22 | ABP30300 | Human brain expres |
| 734 | 21 | 22 | ABG63534 | Peptide #2984 enco | 807 | 6 | 0.8 | 22 | ABP30300 | Human bone marrow  |
| 735 | 21 | 22 | ABG63535 | Protein #2915 enco | 808 | 6 | 0.8 | 22 | ABP30300 | Peptide #1998 enco |
| 736 | 21 | 22 | ABG63536 | Human brain expres | 809 | 6 | 0.8 | 22 | ABP30300 | Peptide #2091 enco |
| 737 | 21 | 22 | ABG63537 | Human bone marrow  | 810 | 6 | 0.8 | 22 | ABP30300 |                    |
| 738 | 21 | 22 | ABG63538 | Peptide #2927 enco | 811 | 6 | 0.8 | 22 | ABP30300 |                    |
| 739 | 21 | 22 | ABG63539 |                    | 812 | 6 | 0.8 | 22 | ABP30300 |                    |

|     |   |     |    |    |          |                      |     |    |    |          |                    |
|-----|---|-----|----|----|----------|----------------------|-----|----|----|----------|--------------------|
| 813 | 6 | 0.8 | 52 | 22 | AAW03308 | Peptide #1990 enco   | 886 | 63 | 22 | AAW23858 | Human EST encoded  |
| 814 | 6 | 0.8 | 52 | 22 | ABP04620 | Human ORFX protein   | 887 | 63 | 22 | AAW24228 | Human EST encoded  |
| 815 | 6 | 0.8 | 52 | 24 | ABP79220 | N. gonorrhoeae ami   | 888 | 64 | 22 | AU41697  | Propionibacterium  |
| 816 | 6 | 0.8 | 53 | 22 | ABP30061 | Peptide #2712 enco   | 889 | 64 | 22 | AAW38525 | Peptide #12562 enc |
| 817 | 6 | 0.8 | 53 | 22 | ABP35232 | Peptide #2738 enco   | 890 | 64 | 22 | ABG47287 | Human peptide enco |
| 818 | 6 | 0.8 | 53 | 22 | ABP20672 | Protein #2671 enco   | 891 | 64 | 23 | ABG60030 | Human DTHP polye   |
| 819 | 6 | 0.8 | 53 | 22 | AAW03980 | Peptide #2662 enco   | 892 | 65 | 13 | AAW22135 | PDGI subunit b. H  |
| 820 | 6 | 0.8 | 53 | 22 | ABG38011 | Human peptide enco   | 893 | 65 | 21 | AAW04147 | Arabidopsis thalia |
| 821 | 6 | 0.8 | 53 | 21 | ABG54014 | Human pancreatic c   | 894 | 65 | 21 | AAW23398 | Arabidopsis thalia |
| 822 | 6 | 0.8 | 54 | 21 | AAW27638 | Human secreted pro   | 895 | 65 | 23 | ABP40519 | Staphylococcus epi |
| 823 | 6 | 0.8 | 54 | 21 | AAW40015 | Propionibacterium    | 896 | 65 | 24 | ABF78492 | N. gonorrhoeae ami |
| 824 | 6 | 0.8 | 54 | 22 | AAW85640 | Human immune/haema   | 897 | 66 | 22 | AAU53716 | Propionibacterium  |
| 825 | 6 | 0.8 | 54 | 22 | ABP05936 | Human ORFX protein   | 898 | 66 | 23 | ABP00425 | Human ORFX protein |
| 826 | 6 | 0.8 | 54 | 23 | ABG00527 | Human secreted pro   | 899 | 66 | 21 | AAW41347 | Human ORFX ORF111  |
| 827 | 6 | 0.8 | 55 | 21 | ABG17988 | Novel human diagno   | 900 | 67 | 22 | ABW70603 | Drosophila melanog |
| 828 | 6 | 0.8 | 55 | 23 | ABP32280 | Human ORF1253 prot   | 901 | 67 | 22 | AAU43607 | Propionibacterium  |
| 829 | 6 | 0.8 | 55 | 23 | ABP33510 | Human ORF2483 prot   | 902 | 67 | 22 | AAW88961 | Human immune/haema |
| 830 | 6 | 0.8 | 55 | 23 | ABP76158 | Human GENSER prote   | 903 | 67 | 22 | AAW00968 | Human polypeptide  |
| 831 | 6 | 0.8 | 55 | 24 | ABP76159 | Human GENSER prote   | 904 | 67 | 23 | ABF09708 | Human ORFX protein |
| 832 | 6 | 0.8 | 56 | 21 | AAW37684 | Arabidopsis thalia   | 905 | 68 | 22 | AAU22720 | Human prostate can |
| 833 | 6 | 0.8 | 56 | 22 | AAW50271 | Propionibacterium    | 906 | 68 | 22 | AAW94692 | Human reproductive |
| 834 | 6 | 0.8 | 56 | 22 | AAW67464 | Propionibacterium    | 907 | 68 | 22 | AAW38937 | Human reproduction |
| 835 | 6 | 0.8 | 56 | 22 | AAW08610 | Human polypeptide    | 908 | 69 | 14 | AAW26135 | LD78 Lys60>Gln, As |
| 836 | 6 | 0.8 | 56 | 22 | AAW62171 | Human gene 16-enco   | 909 | 69 | 21 | AAW16396 | Adenovirus mutated |
| 837 | 6 | 0.8 | 56 | 22 | AAW62208 | Human gene 16-enco   | 910 | 69 | 21 | AAW16396 | Arabidopsis thalia |
| 838 | 6 | 0.8 | 56 | 22 | ABG63566 | Human albumin fusi   | 911 | 69 | 21 | AAW16396 | Arabidopsis thalia |
| 839 | 6 | 0.8 | 56 | 23 | ABG63566 | Human albumin fusi   | 912 | 69 | 21 | AAW16396 | Arabidopsis thalia |
| 840 | 6 | 0.8 | 56 | 23 | ABP00983 | Human ORFX protein   | 913 | 69 | 22 | AAU44795 | Arabidopsis thalia |
| 841 | 6 | 0.8 | 57 | 22 | ABG51794 | Human liver peptid   | 914 | 69 | 23 | ABP39313 | Propionibacterium  |
| 842 | 6 | 0.8 | 57 | 22 | AAU47111 | Propionibacterium    | 915 | 70 | 23 | ABP08564 | Staphylococcus epi |
| 843 | 6 | 0.8 | 57 | 22 | AAW42507 | Propionibacterium    | 916 | 70 | 23 | AAW23068 | Human ORFX protein |
| 844 | 6 | 0.8 | 57 | 22 | ABG28476 | Novel human diagno   | 917 | 71 | 18 | AAW23068 | Canine IGE heavy c |
| 845 | 6 | 0.8 | 57 | 22 | AAW23883 | Human EST encoded    | 918 | 71 | 22 | ABG59686 | Human liver peptid |
| 846 | 6 | 0.8 | 57 | 22 | AAW62742 | Murine secreted al   | 919 | 71 | 22 | ABW44318 | Peptide #11824 enc |
| 847 | 6 | 0.8 | 57 | 23 | ABP31099 | Human ORF72 protei   | 920 | 71 | 22 | AAW65356 | Protein #9176 enco |
| 848 | 6 | 0.8 | 58 | 22 | ABG06251 | Novel human diagno   | 921 | 71 | 22 | AAW78049 | Human brain expre  |
| 849 | 6 | 0.8 | 58 | 22 | AAU22112 | Human cardiovascular | 922 | 71 | 22 | AAW11790 | Human bone marrow  |
| 850 | 6 | 0.8 | 59 | 22 | ABG59327 | Human liver peptid   | 923 | 71 | 22 | AAW21942 | Human polypeptide  |
| 851 | 6 | 0.8 | 59 | 22 | ABW42919 | Peptide #10425 enc   | 924 | 71 | 23 | ABG47063 | Peptide #8176 enco |
| 852 | 6 | 0.8 | 59 | 22 | ABW26169 | Protein #8168 enco   | 925 | 71 | 23 | ABG09521 | Human Reptide enco |
| 853 | 6 | 0.8 | 59 | 22 | AAU32256 | Novel human secret   | 926 | 72 | 22 | AAU20729 | Human novel foetal |
| 854 | 6 | 0.8 | 59 | 22 | AAW63820 | Human brain expre    | 927 | 72 | 22 | ABG69626 | Human secreted pro |
| 855 | 6 | 0.8 | 59 | 22 | AAW76634 | Human bone marrow    | 928 | 72 | 23 | AAW39660 | Propionibacterium  |
| 856 | 6 | 0.8 | 59 | 22 | AAW00097 | Human polypeptide    | 929 | 73 | 22 | ABW80911 | Murine eotaxin mat |
| 857 | 6 | 0.8 | 59 | 22 | AAW36739 | Peptide #10776 enc   | 930 | 73 | 23 | ABW80911 | M. tuberculosis an |
| 858 | 6 | 0.8 | 59 | 22 | AAW36739 | Amino acid sequenc   | 931 | 74 | 20 | AAW39213 | M. tuberculosis re |
| 859 | 6 | 0.8 | 59 | 23 | ABG63230 | Human prostate spe   | 932 | 74 | 20 | AAW39070 | Novel human diagno |
| 860 | 6 | 0.8 | 60 | 21 | AAW21451 | Cone snail alpha-c   | 933 | 74 | 22 | ABG15229 | Novel human diagno |
| 861 | 6 | 0.8 | 60 | 22 | ABG56511 | Human liver peptid   | 934 | 74 | 22 | ABG28523 | Human polypeptide  |
| 862 | 6 | 0.8 | 60 | 22 | ABG16809 | Novel human diagno   | 935 | 74 | 22 | AAW10827 | Human polypeptide  |
| 863 | 6 | 0.8 | 60 | 22 | ABG18662 | Novel human diagno   | 936 | 74 | 24 | ABW06307 | Maize GBSS glycosy |
| 864 | 6 | 0.8 | 60 | 22 | ABW41068 | Peptide #8574 enco   | 937 | 75 | 21 | AAW43404 | Human cancer assoc |
| 865 | 6 | 0.8 | 60 | 22 | ABW21333 | Protein #7132 enco   | 938 | 75 | 21 | AAW43404 | Arabidopsis thalia |
| 866 | 6 | 0.8 | 60 | 22 | AAW61926 | Human brain expre    | 939 | 75 | 22 | ABW44202 | Propionibacterium  |
| 867 | 6 | 0.8 | 60 | 22 | AAW74728 | Human bone marrow    | 940 | 75 | 22 | AAW39741 | Drosophila melanog |
| 868 | 6 | 0.8 | 60 | 22 | AAW01895 | Human polypeptide    | 941 | 75 | 23 | ABF32201 | Propionibacterium  |
| 869 | 6 | 0.8 | 60 | 22 | AAW20394 | Peptide #6828 enco   | 942 | 75 | 23 | ABF05338 | Human ORF1174 prot |
| 870 | 6 | 0.8 | 60 | 22 | AAW34845 | Peptide #8892 enco   | 943 | 76 | 21 | ABW44909 | Human ORFX protein |
| 871 | 6 | 0.8 | 60 | 22 | ABW68685 | Human TGPbetal pro   | 944 | 76 | 21 | AAW01406 | Human secreted pro |
| 872 | 6 | 0.8 | 60 | 23 | ABG44532 | Human peptide enco   | 945 | 76 | 21 | ABG16681 | Human secreted pro |
| 873 | 6 | 0.8 | 60 | 23 | ABP36512 | Human ORF1588 prot   | 946 | 76 | 22 | ABW03650 | Novel human diagno |
| 874 | 6 | 0.8 | 60 | 23 | ABP09499 | Human ORFX protein   | 947 | 76 | 22 | ABW03650 | Human musculocele  |
| 875 | 6 | 0.8 | 61 | 22 | ABG16951 | Novel human diagno   | 948 | 76 | 22 | AAW92722 | Human pancreatic c |
| 876 | 6 | 0.8 | 61 | 23 | ABG40374 | Human peptide enco   | 949 | 76 | 23 | AAW92722 | Human digestive sy |
| 877 | 6 | 0.8 | 62 | 18 | AAW30331 | Fragment of growth   | 950 | 76 | 23 | ABW51522 | Helicobacter pylor |
| 878 | 6 | 0.8 | 62 | 22 | AAW86507 | Novel human connec   | 951 | 76 | 24 | ABU12944 | Staphylococcus epi |
| 879 | 6 | 0.8 | 62 | 22 | AAW86507 | Human gene 8 enco    | 952 | 76 | 24 | ABW68810 | Novel human muscu  |
| 880 | 6 | 0.8 | 62 | 22 | AAW86507 | Human albumin fusi   | 953 | 77 | 22 | ABW68810 | Novel human muscu  |
| 881 | 6 | 0.8 | 62 | 23 | ABG63525 | Human albumin fusi   | 954 | 77 | 22 | AAW53404 | Propionibacterium  |
| 882 | 6 | 0.8 | 62 | 23 | ABG65032 | Human albumin fusi   | 955 | 77 | 22 | AAW53404 | Propionibacterium  |
| 883 | 6 | 0.8 | 62 | 23 | ABG65032 | Human secreted pro   | 956 | 77 | 24 | AAU00199 | Human novel polye  |
| 884 | 6 | 0.8 | 63 | 21 | AAW00842 | Human secreted pro   | 957 | 78 | 22 | AAU42924 | Propionibacterium  |
| 885 | 6 | 0.8 | 63 | 22 | AAW59103 | Propionibacterium    | 958 | 78 | 22 | AAU22637 | Novel human colon  |
|     | 6 | 0.8 | 63 | 22 | AAW02709 | Human polypeptide    | 959 | 78 | 22 | AAW92643 | Human digestive sy |
|     | 6 | 0.8 | 63 | 22 | AAW02709 | Human polypeptide    | 958 | 78 | 22 | AAW92643 | S cerevisiae apopt |

|      |   |     |    |    |          |                     |
|------|---|-----|----|----|----------|---------------------|
| 959  | 6 | 0.8 | 79 | 21 | AAG29579 | Arabidopsis thalia  |
| 960  | 6 | 0.8 | 79 | 22 | ABG49091 | Human liver peptid  |
| 961  | 6 | 0.8 | 79 | 22 | ABB29091 | Peptide #1742 enco  |
| 962  | 6 | 0.8 | 79 | 22 | ABB34252 | Peptide #1758 enco  |
| 963  | 6 | 0.8 | 79 | 22 | ABB19688 | Protein #1687 enco  |
| 964  | 6 | 0.8 | 79 | 22 | AAM55045 | Human brain expres  |
| 965  | 6 | 0.8 | 79 | 22 | AAM67435 | Human bone marrow   |
| 966  | 6 | 0.8 | 79 | 22 | AAM15261 | Peptide #1695 enco  |
| 967  | 6 | 0.8 | 79 | 22 | AAM27726 | Peptide #1763 enco  |
| 968  | 6 | 0.8 | 79 | 22 | AAM30306 | Peptide #1688 enco  |
| 969  | 6 | 0.8 | 79 | 22 | AAQ74217 | Human colon cancer  |
| 970  | 6 | 0.8 | 79 | 23 | ABU51608 | Helicobacter pylori |
| 971  | 6 | 0.8 | 79 | 23 | ABG37058 | Human peptide enco  |
| 972  | 6 | 0.8 | 79 | 23 | ABP06059 | Human ORFX protein  |
| 973  | 6 | 0.8 | 80 | 21 | AAG50015 | Arabidopsis thalia  |
| 974  | 6 | 0.8 | 80 | 22 | ABG50309 | Human liver peptid  |
| 975  | 6 | 0.8 | 80 | 22 | ABG55870 | Human liver peptid  |
| 976  | 6 | 0.8 | 80 | 22 | AAJ86561 | Novel human connec  |
| 977  | 6 | 0.8 | 80 | 22 | ABG30282 | Peptide #2933 enco  |
| 978  | 6 | 0.8 | 80 | 22 | ABG40534 | Peptide #8040 enco  |
| 979  | 6 | 0.8 | 80 | 22 | ABB15177 | Human nervous syst  |
| 980  | 6 | 0.8 | 80 | 22 | AAJ74107 | Human bone marrow   |
| 981  | 6 | 0.8 | 80 | 22 | AAO01706 | Human polypeptide   |
| 982  | 6 | 0.8 | 80 | 22 | AAW28955 | Peptide #2992 enco  |
| 983  | 6 | 0.8 | 80 | 22 | AAV34267 | Peptide #8104 enco  |
| 984  | 6 | 0.8 | 80 | 22 | AAU19860 | Human novel extrac  |
| 985  | 6 | 0.8 | 80 | 23 | ABG44013 | Human peptide enco  |
| 986  | 6 | 0.8 | 80 | 23 | ABP48080 | Human polypeptide   |
| 987  | 6 | 0.8 | 81 | 21 | AAE69334 | HIV-1 non-subtype   |
| 988  | 6 | 0.8 | 81 | 21 | AAE69339 | HIV-1 non-subtype   |
| 989  | 6 | 0.8 | 81 | 22 | AAU57655 | Propionibacterium   |
| 990  | 6 | 0.8 | 81 | 23 | ABG69884 | Human secretory pr  |
| 991  | 6 | 0.8 | 82 | 20 | AAV23877 | Winter flounder an  |
| 992  | 6 | 0.8 | 82 | 21 | AAJ17117 | Arabidopsis thalia  |
| 993  | 6 | 0.8 | 82 | 21 | AAJ19327 | Arabidopsis thalia  |
| 994  | 6 | 0.8 | 82 | 22 | ABG05861 | Novel human diagno  |
| 995  | 6 | 0.8 | 82 | 23 | ABP31214 | Human ORF187 prote  |
| 996  | 6 | 0.8 | 83 | 18 | AAJ19092 | Trypanosoma cruzi   |
| 997  | 6 | 0.8 | 83 | 20 | AAV23318 | Peptide containing  |
| 998  | 6 | 0.8 | 83 | 21 | AAJ29578 | Arabidopsis thalia  |
| 999  | 6 | 0.8 | 83 | 22 | ABP58839 | Human testicular a  |
| 1000 | 6 | 0.8 | 83 | 22 | ABG05602 | Novel human diagno  |

ALIGNMENTS

|          |  |
|----------|--|
| RESULT 1 |  |
| AAV53895 |  |
| ID       | AAV53895 standard; Protein; 769 AA.                            |
| XX       |  |
| AC       | AAV53895;  |
| XX       |  |
| DT       | 13-MAR-2000 (first entry)                                      |
| XX       |  |
| DE       | A Neisseria meningitidis antigenic protein designated BASB030. |
| XX       |  |
| KW       | Antigenic polypeptide; BASB030; serotype B strain ATCC 13090;  |
| KW       | vaccine; infection; bacteremia; meningitis.                    |
| XX       |  |
| OS       | Neisseria meningitidis.  |
| XX       |  |
| PN       | WO9961620-A2.  |
| XX       |  |
| PD       | 02-DEC-1999.   |
| XX       |  |
| PF       | 26-MAY-1999; 99WO-EP03603.                                     |
| XX       |  |
| PR       | 26-MAY-1998; 98GB-0011260.                                     |
| XX       |  |
| PA       | (SMK ) SMITHKLINE BEECHAM BIOLOGICALS.                         |
| XX       |  |
| PI       | Ruelle J;  |

|                           |   |
|---------------------------|---|
| XX                        | WPI; 2000-072624/06.  |
| DR                        | N-PSDB; AAZ36857.   |
| XX                        |   |
| PT                        | New isolated Neisseria meningitidis polypeptides and polynucleotides, used to develop products for the diagnosis, prevention and treatment of infections  |
| PT                        |   |
| XX                        | Claim 5; Page 86-88; 97pp; English.   |
| PS                        |   |
| XX                        | The present sequence represents a Neisseria meningitidis antigenic polypeptide, designated BASB030. It was identified from N. meningitidis serotype B strain ATCC 13090. The nucleotide sequence was first identified in the Incyte Pathoseq database of unfinished genomic DNA sequences of this strain. The polypeptides or polynucleotides can be used in vaccine compositions for preventing N. meningitidis infections, e.g. bacteremia and meningitis. The antibodies against the protein can be used for treating N. meningitidis disease. The products can also be used for diagnosis of disease, staging of disease or response of an infectious organism to drugs, as well as for drug screening. |
| CC                        |   |
| XX                        | Sequence 769 AA;  |
| SQ                        |   |
| Query Match               | 100.0%; Score 769; DB 21; Length 769;   |
| Best Local Similarity     | 100.0%; Pred. No. 0;  |
| Matches 769; Conservative | 0; Mismatches 0; Indels 0; Gaps 0;  |
| QY                        | 1 MNTKTKIISGLFVATAPOTASAGNITDIKVSILPNKQIKVKVSFDEKIVNPTGFTVSS 60   |
| DB                        | 1 MNTKTKIISGLFVATAPOTASAGNITDIKVSILPNKQIKVKVSFDEKIVNPTGFTVSS 60   |
| QY                        | 61 PARIALDPEQTGISMDQVLEYADPFLSKISAQNSSPARLVNLNPKPGQYNTVEVRGNKV 120  |
| DB                        | 61 PARIALDPEQTGISMDQVLEYADPFLSKISAQNSSPARLVNLNPKPGQYNTVEVRGNKV 120  |
| QY                        | 121 WIFINESDDTVSAPARPAKPAKPAKQCGRTVYQVRSIRIOTLYPGKTTAAAPFTES 180  |
| DB                        | 121 WIFINESDDTVSAPARPAKPAKPAKQCGRTVYQVRSIRIOTLYPGKTTAAAPFTES 180  |
| QY                        | 181 VVVSAPSPAKQQAASAKQQTAAAPAKQQAAPAKQTNIDFRKGNKAGII 240  |
| DB                        | 181 VVVSAPSPAKQQAASAKQQTAAAPAKQQAAPAKQTNIDFRKGNKAGII 240  |
| QY                        | 241 ELAALGFAGQPDISOQHDHIIIVTLKXNHTLPTLQSLVDVDFKTPVQKVLKLNNDTQL 300  |
| DB                        | 241 ELAALGFAGQPDISOQHDHIIIVTLKXNHTLPTLQSLVDVDFKTPVQKVLKLNNDTQL 300  |
| QY                        | 301 IITTAGNWLNVKNSAAGVYFTFQVLPKQNLBSGGVNNAPKFTTGRKISLDFQDVEIRTI 360   |
| DB                        | 301 IITTAGNWLNVKNSAAGVYFTFQVLPKQNLBSGGVNNAPKFTTGRKISLDFQDVEIRTI 360   |
| QY                        | 361 LQILAKESGMNIVASDSVNGKMTLSLKDVDPDQALDLMQARNLDMROQGNIVNIAPRDE 420   |
| DB                        | 361 LQILAKESGMNIVASDSVNGKMTLSLKDVDPDQALDLMQARNLDMROQGNIVNIAPRDE 420   |
| QY                        | 421 LIAKOKAFLOAKDIADLGALYSQNFQKYNVEEFRSILRLDNADTTGNRNTLVSGRGS 480   |
| DB                        | 421 LIAKOKAFLOAKDIADLGALYSQNFQKYNVEEFRSILRLDNADTTGNRNTLVSGRGS 480   |
| QY                        | 481 VLIDPATNTLIVTDTSRVIEKFKLIDELDPAQQVMIEARIVEAADGFSRDLGVKFGAT 540  |
| DB                        | 481 VLIDPATNTLIVTDTSRVIEKFKLIDELDPAQQVMIEARIVEAADGFSRDLGVKFGAT 540  |
| QY                        | 541 GKXKLKNDTSFAGWGVNSGFGGDDKWAETKINLPITAAANSISLVRAISSGALNELSA 600  |
| DB                        | 541 GKXKLKNDTSFAGWGVNSGFGGDDKWAETKINLPITAAANSISLVRAISSGALNELSA 600  |
| QY                        | 601 SESLSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANGSSSTNTTELKAVLGLTVTP 660  |
| DB                        | 601 SESLSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANGSSSTNTTELKAVLGLTVTP 660  |
| QY                        | 661 NITPDQOIIMTVKINKDSPAQCASGNQITLICSTKNLNTQAMVENGGLTVGIIYEEDNG 720   |

```

Db      661 NTPDGGIIMTVKINKDSPAQCASGNQITLCISTKNLNTQAMVNGGTLIVGGIYEEDNG 720
Qy      721 NLTLPVLLGDIPIVGNLFKTRGKKTDRRELLIFITPRIMGTAGNSLRY 769
Db      721 NLTLPVLLGDIPIVGNLFKTRGKKTDRRELLIFITPRIMGTAGNSLRY 769

RESULT 2
AA53896
ID  AA53896 standard; Protein; 769 AA.
XX
XX  AA53896;
DT  13-MAR-2000 (first entry)
DE  A Neisseria meningitidis antigenic protein designated BASB030.
XX
XX  Antigenic polypeptide; BASB030; serotype B strain ATCC 13090;
KW  vaccine; infection; bacteremia; meningitis.
XX
XX  Neisseria meningitidis.
OS
XX  WO9961620-A2.
PN
PD  02-DEC-1999.
PF
XX  26-MAY-1999; 99WO-EP03603.
XX
XX  26-MAY-1998; 98GB-0011260.
PA  (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX  Ruelle J;
XX
XX  WPI; 2000-072624/06.
DR  N-PSDB; AA236858.
XX
XX  New isolated Neisseria meningitidis polypeptides and polynucleotides,
PT  used to develop products for the diagnosis, prevention and treatment of
PT  infections -
XX
XX  Claim 3; Page 90-92; 97pp; English.
XX
XX  The present sequence represents a Neisseria meningitidis antigenic
CC  polypeptide, designated BASB030. It was identified from N. meningitidis
CC  serotype B strain ATCC 13090. The nucleotide sequence was first
CC  identified in the Incyte Pathoseq database of unfinished genomic DNA
CC  sequences of this strain. The polypeptides or polynucleotides can be
CC  used in vaccine compositions for preventing N. meningitidis infections,
CC  e.g. bacteremia and meningitis. The antibodies against the protein can
CC  be used for treating N. meningitidis disease. The products can also
CC  be used for diagnosis of disease, staging of disease or response of an
CC  infectious organism to drugs, as well as for drug screening.
XX
XX  Sequence 769 AA;
XX

Query Match 77.6%; Score 597; DB 21; Length 769;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      173 AAAPFTESVVSAPFSPAKQQAASAKQQTAAAPAKQQTAAAPAKQQTIDFRK 232
Db      173 AAAPFTESVVSAPFSPAKQQAASAKQQTAAAPAKQQTAAAPAKQQTIDFRK 232

Qy      233 DGNVAGIIEALGFAGQPDISOQHDHIIIVTKNHTLPTTLORSLDVADFPTPVOKVTLK 292
Db      233 DGNVAGIIEALGFAGQPDISOQHDHIIIVTKNHTLPTTLORSLDVADFPTPVOKVTLK 292

Qy      293 RLNNDTQLIITTAGNWLKNSAAGPYFTFQVLPKKQNLKESGVNNAKPTFTGRKISLDF 352
Db      293 RLNNDTQLIITTAGNWLKNSAAGPYFTFQVLPKKQNLKESGVNNAKPTFTGRKISLDF 352

Qy      353 QDVEIRTIQLAKESGMNIVASDSVNGKMTLSLKDVPHDQALDLVMAARNLDMRQQNI 412

```

```

Db      353 QDVEIRTIQLAKESGMNIVASDSVNGKMTLSLKDVPHDQALDLVMAARNLDMRQQNI 412
Qy      413 VNIAPRDELLAKDKAFLOAEKDIADLGALYSQNFOLKYKXVEFRSILRLDNADTTGNRN 472
Db      413 VNIAPRDELLAKDKAFLOAEKDIADLGALYSQNFOLKYKXVEFRSILRLDNADTTGNRN 472

Qy      473 TLVSGRGSVLIDPATNTLIVTDRSVIEKFKLIDELVPAQVMIEARIVEADGFSRD 532
Db      473 TLVSGRGSVLIDPATNTLIVTDRSVIEKFKLIDELVPAQVMIEARIVEADGFSRD 532

Qy      533 LGVKFGATGKKLKNKNTSAFGWGVNSGFGGDDKWAETKINLPITAAANSISLVRAISSG 592
Db      533 LGVKFGATGKKLKNKNTSAFGWGVNSGFGGDDKWAETKINLPITAAANSISLVRAISSG 592

Qy      593 ALNLELSASESLSKTTLANPRVLITQNRKAEKIESGYEIPFTVTSIANGSSNTTELKKA 652
Db      593 ALNLELSASESLSKTTLANPRVLITQNRKAEKIESGYEIPFTVTSIANGSSNTTELKKA 652

Qy      653 VLGLTVPNTIPDQIIMTVKINKDSPAQCASGNQITLCISTKNLNTQAMVNGGTLIVG 712
Db      653 VLGLTVPNTIPDQIIMTVKINKDSPAQCASGNQITLCISTKNLNTQAMVNGGTLIVG 712

Qy      713 GIYEEDNGNTLTKVPLLLGDIPIVGNLFKTRGKKTDRRELLIFITPRIMGTAGNSLRY 769
Db      713 GIYEEDNGNTLTKVPLLLGDIPIVGNLFKTRGKKTDRRELLIFITPRIMGTAGNSLRY 769

RESULT 3
AA53897
ID  AA53897 standard; Protein; 769 AA.
XX
XX  AA53897;
DT  13-MAR-2000 (first entry)
DE  A Neisseria meningitidis antigenic protein designated BASB030.
XX
XX  Antigenic polypeptide; BASB030; serotype B strain H44/76;
KW  vaccine; infection; bacteremia; meningitis.
XX
XX  Neisseria meningitidis.
OS
XX  WO9961620-A2.
PN
PD  02-DEC-1999.
PF
XX  26-MAY-1999; 99WO-EP03603.
XX
XX  26-MAY-1998; 98GB-0011260.
PA  (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX  Ruelle J;
XX
XX  WPI; 2000-072624/06.
DR  N-PSDB; AA236859.
XX
XX  New isolated Neisseria meningitidis polypeptides and polynucleotides,
PT  used to develop products for the diagnosis, prevention and treatment of
PT  infections -
XX
XX  Claim 3; Page 93-96; 97pp; English.
XX

The present sequence represents a Neisseria meningitidis antigenic
CC  polypeptide, designated BASB030. It was identified from N. meningitidis
CC  serotype B strain H44/76. The nucleotide sequence was first
CC  identified in the Incyte Pathoseq database of unfinished genomic DNA
CC  sequences of this strain. The polypeptides or polynucleotides can be
CC  used in vaccine compositions for preventing N. meningitidis infections,
CC  e.g. bacteremia and meningitis. The antibodies against the protein can
CC  be used for treating N. meningitidis disease. The products can also
CC  be used for diagnosis of disease, staging of disease or response of an

```

CC infectious organism to drugs, as well as for drug screening.

XX SQ Sequence 769 AA; Query Match 46.3%; Score 356; DB 21; Length 769; Best Local Similarity 99.6%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0; Matches 556; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 212 AAPAKQAAAPAKQTNIDFRKDGKNAGIIEALGAFAGQPDISQOHDHIIIVTLKHHTLPT 271  
DB 212 AAPAKQAAAPAKQTNIDFRKDGKNAGIIEALGAFAGQPDISQOHDHIIIVTLKHHTLPT 271  
QY 272 TLORSIDVADFKTPVKQVTLKRLNNDTQLIIITAGNWLWNKSAAPGYFTFQVLPKKQNL 331  
DB 272 TLORSIDVADFKTPVKQVTLKRLNNDTQLIIITAGNWLWNKSAAPGYFTFQVLPKKQNL 331  
QY 332 ESGVNNAPKTFTRKISLDFQVIEIRTIQILAKESGMNIVASDSVNGKMTLSLKDVPW 391  
DB 332 ESGVNNAPKTFTRKISLDFQVIEIRTIQILAKESGMNIVASDSVNGKMTLSLKDVPW 391  
QY 392 DOALDLYMOARNLDMROQGNIVNIAPRDELLAKDKAFLOAEKDIAADLGALYSONFOLKYK 451  
DB 392 DOALDLYMOARNLDMROQGNIVNIAPRDELLAKDKAFLOAEKDIAADLGALYSONFOLKYK 451  
QY 452 NVBEFRSILRLDNADTTGNRNTLVSGRGSVLIDPATNTLIVTDRSVIEKFRKLIDELDV 511  
DB 452 NVBEFRSILRLDNADTTGNRNTLVSGRGSVLIDPATNTLIVTDRSVIEKFRKLIDELDV 511  
QY 512 PAQOVMIARIVEAADGFSRLGVKFGATGKKLNDTSAGFGVNSGFGDDDKWGAETK 571  
DB 512 PAQOVMIARIVEAADGFSRLGVKFGATGKKLNDTSAGFGVNSGFGDDDKWGAETK 571  
QY 572 INLPITAAANSISLVRAISSGALNLELSASESLSKTKTLANPRVLTQNRKEAKIESGYEI 631  
DB 572 INLPITAAANSISLVRAISSGALNLELSASESLSKTKTLANPRVLTQNRKEAKIESGYEI 631  
QY 632 PFTVTSIANGSSNTTELKXAVLGLTVPNTIPDQQLIMTVKINKDSPACASGNQTLIC 691  
DB 632 PFTVTSIANGSSNTTELKXAVLGLTVPNTIPDQQLIMTVKINKDSPACASGNQTLIC 691  
QY 692 ISTKNLNTQAAVNGGTLIVGGIYEEDNGNTLTKVPLLDGIPVIGNLFKTRGKKTDRREL 751  
DB 692 ISTKNLNTQAAVNGGTLIVGGIYEEDNGNTLTKVPLLDGIPVIGNLFKTRGKKTDRREL 751  
QY 752 LIFITPRIMGTAGNSLRY 769  
DB 752 LIFITPRIMGTAGNSLRY 769

RESULT 4  
ABP77693 ID ABP77693 standard; Protein; 723 AA.

XX AC ABP77693;  
XX DT 07-MAR-2003 (first entry)  
XX DE N. gonorrhoeae amino acid sequence SEQ ID 1916.  
XX KW Antibacterial; infection; vaccine; gene therapy.

XX OS Neisseria gonorrhoeae.

XX PN WO200279243-A2.

XX PD 10-OCT-2002.

XX PF 12-FEB-2002; 2002WO-IB02069.

XX PR 12-FEB-2001; 2001GB-0003424.

XX PA (CHIR-) CHIRON SPA.

XX XX

PI Fontana MR, Pizza M, Masignani V, Monaci E;  
XX WPI; 2003-058415/05.  
DR N-PSDB; ABZ38663.  
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
PT medicament for treating or preventing N. gonorrhoeae infection -  
XX Disclousure; Page 336; 815pp; English.  
PS The present invention relates to proteins from Neisseria gonorrhoeae.  
XX Also disclosed are the nucleic acid molecules encoding the proteins and  
CC antibodies that specifically bind to the proteins. The composition  
CC comprising the protein, nucleic acid or antibody is useful for the  
CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
CC infection, this may be in the form of a vaccine or gene therapy.  
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
CC molecules of the invention.  
XX Sequence 723 AA;

QY Query Match 21.1%; Score 162; DB 24; Length 723;  
DB Best Local Similarity 100.0%; Pred. No. 1.3e-152;  
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 360 GKMVLSLKVDPWDQALDLYMOARNLDMROQGNIVNIAPRDELLAKDKAFLOAEKDIAADLG 439  
DB 334 GKMVLSLKVDPWDQALDLYMOARNLDMROQGNIVNIAPRDELLAKDKAFLOAEKDIAADLG 393  
QY 440 ALYSQNFOLKYKNVEFRSILRLDNADTTGNRNTLVSGRGSVLIDPATNTLIVTDRSVI 499  
DB 394 ALYSQNFOLKYKNVEFRSILRLDNADTTGNRNTLVSGRGSVLIDPATNTLIVTDRSVI 453  
QY 500 EXFRKLIDELVPAQOVMIARIVEAADGFSRLGVKFGATG 541  
DB 454 EXFRKLIDELVPAQOVMIARIVEAADGFSRLGVKFGATG 495

RESULT 5

ABG91062 ID ABG91062 standard; Protein; 720 AA.

XX AC ABG91062;  
XX DT 29-NOV-2002 (first entry)  
XX DE Neisseria gonorrhoeae pilus biogenesis protein #3.  
XX KW Gram-negative bacterial bleb; PorB; outer membrane protein;  
XX Chlamydia trachomatis infection; Chlamydia pneumoniae infection;  
XX KW protective antigen; antibacterial; vaccine.

XX OS Neisseria gonorrhoeae.

XX PN WO200262380-A2.

XX PD 15-AUG-2002.

XX PF 08-FEB-2002; 2002WO-EP01356.

XX PR 08-FEB-2001; 2001GB-0003169.

XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Berthet FJ, Lobet Y, Poolman J, Verlant VGCL;

XX WPI; 2002-657510/70.

XX DR N-PSDB; ABS67380.

XX Novel gram-negative bacterial bleb presenting on its surface porB outer  
PT membrane protein from Chlamydia trachomatis or protective antigen from  
PT Chlamydia pneumoniae, useful for preventing Chlamydia infection -

XX XX



PS Disclosure; Page 52; 75pp; English.

XX The present invention relates to a new gram-negative bacterial bleb

CC presenting on its surface the PorB outer membrane protein from Chlamydia

CC trachomatis, or a protective antigen from C. pneumoniae. The invention

CC is useful for preventing C. trachomatis or C. pneumoniae infection in a

CC host. The present amino acid sequence represents a Neisseria gonorrhoeae

CC protein as described in the invention.

XX

XX Sequence 720 AA;

SQ

Query Match 19.5%; Score 150; DB 23; Length 720;

Best Local Similarity 100.0%; Pred. No. 1.3e-140; Indels 0; Gaps 0;

Matches 150; Conservative 0; Mismatches 0;

QY 1 MNTKLTIIISGLFVATAAFQASAGNITDIKVSSLPNQKIVKVSFDEIIVNPTGFTVSS 60

DB 1 MNTKLTIIISGLFVATAAFQASAGNITDIKVSSLPNQKIVKVSFDEIIVNPTGFTVSS 60

QY 61 PARIALDFEQTGISDQVLEFVADPILLSKISAAQNSSRRLVNLNKPQYNTVEVRGNKV 120

DB 61 PARIALDFEQTGISDQVLEFVADPILLSKISAAQNSSRRLVNLNKPQYNTVEVRGNKV 120

QY 121 WFINESDDTVSAPARPAVKAAPAAAPAKQQ 150

DB 121 WFINESDDTVSAPARPAVKAAPAAAPAKQQ 150

RESULT 6

AAV44394

ID AAY44394 standard; Protein; 473 AA.

XX

AC AAY44394;

XX

DT 22-MAR-2000 (first entry)

XX

DE Moraxella catarrhalis BASB031 protein-1, from strain Mc2931(ATCC 43617).

XX

KW BASB031; strain Mc2931; ATCC 43617; PilQ fimbrial assembly protein;

KW diagnosis; treatment; otitis media; sinusitis; pneumonia; screening;

KW homology; nosocomial infection; antibody; ortholog; hybridisation probe.

XX

OS Moraxella catarrhalis.

XX

PN WO9964448-A2.

PD 16-DEC-1999.

XX

PF 31-MAY-1999; 99WO-EP03823.

XX

PR 05-JUN-1998; 98GB-0012163.

XX

PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX

PI Ruelle J, Tommassen JPM, Vinals-Bassols C;

XX

DR WPI; 2000-116523/10.

DR N-PSDB; AA229556.

XX

PT Novel polypeptides used as vaccines for treating Moraxella catarrhalis

PT infections like otitis media and pneumonia -

XX

PS Claim 5; Page 97-98; 121pp; English.

XX

XX The present sequence is the BASB031 protein sequence-1, from strain

CC Mc2931(ATCC 43617), derived from Moraxella catarrhalis. This polypeptide

CC sequence has homology to Pseudomonas aeruginosa, PilQ fimbrial assembly

CC protein. This sequence can be used for prevention and treatment of

CC M.catarrhalis infections, like otitis media, pneumonia, sinusitis and

CC nosocomial infections. The antibodies and polynucleotide sequence can be

CC used for diagnosing infections, staging of disease and for determining

CC the response of an infectious organism to drugs. The DNA sequence is

CC also used as a hybridisation probe in screening process for identifying

CC homologue and orthologs from other species.

XX

SQ Sequence 473 AA;

Query Match 1.2%; Score 9; DB 21; Length 473;

Best Local Similarity 100.0%; Pred. No. 8.4; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 0;

QY 515 QVMIEARIV 523

DB 230 QVMIEARIV 238

RESULT 7

AAV44395

ID AAY44395 standard; Protein; 473 AA.

XX

AC AAY44395;

XX

DT 22-MAR-2000 (first entry)

XX

DE Moraxella catarrhalis BASB031 protein-2, from strain Mc2931(ATCC 43617).

XX

KW BASB031; strain Mc2931; ATCC 43617; PilQ fimbrial assembly protein;

KW diagnosis; treatment; otitis media; sinusitis; pneumonia; screening;

KW homology; nosocomial infection; antibody; ortholog; hybridisation probe.

XX

OS Moraxella catarrhalis.

XX

PN WO9964448-A2.

PD 16-DEC-1999.

XX

PF 31-MAY-1999; 99WO-EP03823.

XX

PR 05-JUN-1998; 98GB-0012163.

XX

PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX

PI Ruelle J, Tommassen JPM, Vinals-Bassols C;

XX

DR WPI; 2000-116523/10.

DR N-PSDB; AA229557.

XX

PT Novel polypeptides used as vaccines for treating Moraxella catarrhalis

PT infections like otitis media and pneumonia -

XX

PS Claim 3; Page 99-101; 121pp; English.

XX

XX The present sequence is the BASB031 protein sequence-2, from strain

CC Mc2931(ATCC 43617), derived from Moraxella catarrhalis. This polypeptide

CC sequence has homology to Pseudomonas aeruginosa, PilQ fimbrial assembly

CC protein. This sequence can be used for prevention and treatment of

CC M.catarrhalis infections, like otitis media, pneumonia, sinusitis and

CC nosocomial infections. The antibodies and polynucleotide sequence can be

CC used for diagnosing infections, staging of disease and for determining

CC the response of an infectious organism to drugs. The DNA sequence is

CC also used as a hybridisation probe in screening process for identifying

CC homologue and orthologs from other species.

XX

SQ Sequence 473 AA;

Query Match 1.2%; Score 9; DB 21; Length 473;

Best Local Similarity 100.0%; Pred. No. 8.4; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 0;

QY 515 QVMIEARIV 523

DB 230 QVMIEARIV 238

RESULT 8

AAV44396

ID AAY44396 standard; Protein; 473 AA.  
XX AC AAY44397;  
XX DT 22-MAR-2000 (first entry)  
XX DE Moraxella catarrhalis BASB031 protein sequence from strain Mc2911.  
XX KW BASB031; strain Mc2911; PilQ fimbrial assembly protein; homology;  
KW diagnosis; treatment; otitis media; pneumonia; sinusitis; antibody;  
KW nosocomial infection; screening; hybridisation probe; ortholog.  
XX OS Moraxella catarrhalis.  
XX PN WO9964448-A2.  
XX PD 16-DEC-1999.  
XX PF 31-MAY-1999; 99WO-EP03823.  
XX PR 05-JUN-1998; 98GB-0012163.  
XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX PI Ruelle J, Tommassen JPM, Vinals-Bassols C;  
XX WPI; 2000-116523/10.  
XX DR N-PSDB; AAZ29558.  
XX PT Novel polypeptides used as vaccines for treating Maraxella catarrhalis  
PT infections like otitis media and pneumonia -  
XX Claim 3; Page 102-103; 121pp; English.  
XX CC The present sequence is the BASB031 polypeptide, from strain Mc2911,  
CC derived from Moraxella catarrhalis. This sequence has homology to  
CC Pseudomonas aeruginosa, PilQ fimbrial assembly protein. This sequence  
CC can be used for prevention and treatment of M.catarrhalis infections.  
CC like otitis media, pneumonia, sinusitis and nosocomial infections. The  
CC antibodies and polynucleotide sequence can be used for diagnosing  
CC infections, staging of disease and for determining the response of an  
CC infectious organism to drugs. The DNA sequence is also used as a  
CC hybridisation probe in screening process for identifying homologue and  
CC orthologs from other species.  
XX SQ Sequence 473 AA;  
XX Query Match 1.2%; Score 9; DB 21; Length 473;  
XX Best Local Similarity 100.0%; Pred.No. 8.4;  
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 515 QVMIEARIV 523  
Db 230 QVMIEARIV 238  
RESULT 9  
ID AAY44397 standard; Protein; 473 AA.  
XX AC AAY44397;  
XX DT 22-MAR-2000 (first entry)  
XX DE Moraxella catarrhalis BASB031 protein sequence from strain Mc2969.  
XX KW BASB031; strain Mc2969; PilQ fimbrial assembly protein; homology;  
KW diagnosis; treatment; otitis media; pneumonia; sinusitis; antibody;  
KW nosocomial infection; screening; hybridisation probe; ortholog.  
XX OS Moraxella catarrhalis.  
XX PN WO9964448-A2.

XX PD 16-DEC-1999.  
XX PF 31-MAY-1999; 99WO-EP03823.  
XX PR 05-JUN-1998; 98GB-0012163.  
XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX PI Ruelle J, Tommassen JPM, Vinals-Bassols C;  
XX WPI; 2000-116523/10.  
XX DR N-PSDB; AAZ29559.  
XX PT Novel polypeptides used as vaccines for treating Maraxella catarrhalis  
PT infections like otitis media and pneumonia -  
XX Claim 3; Page 104-106; 121pp; English.  
XX CC The present sequence is the BASB031 polypeptide, from strain Mc2969,  
CC derived from Moraxella catarrhalis. This sequence has homology to  
CC Pseudomonas aeruginosa, PilQ fimbrial assembly protein. This sequence  
CC can be used for prevention and treatment of M.catarrhalis infections.  
CC like otitis media, pneumonia, sinusitis and nosocomial infections. The  
CC antibodies and polynucleotide sequence can be used for diagnosing  
CC infections, staging of disease and for determining the response of an  
CC infectious organism to drugs. The DNA sequence is also used as a  
CC hybridisation probe in screening process for identifying homologue and  
CC orthologs from other species.  
XX SQ Sequence 473 AA;  
XX Query Match 1.2%; Score 9; DB 21; Length 473;  
XX Best Local Similarity 100.0%; Pred.No. 8.4;  
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 515 QVMIEARIV 523  
Db 230 QVMIEARIV 238  
RESULT 10  
ID AAO17573 standard; Protein; 473 AA.  
XX AC AAO17573;  
XX DT 19-JUL-2002 (first entry)  
XX DE M catarrhalis MCA100992 protein SEQ ID NO: 26.  
XX KW Moraxella; vaccine; respiratory tract infection; antiinflammatory;  
KW auditory; antibacterial; otitis media; sinusitis; pneumonia.  
XX OS Moraxella catarrhalis.  
XX PN WO200218595-A2.  
XX PD 07-MAR-2002.  
XX PF 28-AUG-2001; 2001WO-CA01221.  
XX PR 28-AUG-2000; 2000US-228294P.  
XX PR 28-AUG-2000; 2000US-228295P.  
XX PR 28-AUG-2000; 2000US-228296P.  
XX PR 29-AUG-2000; 2000US-228438P.  
XX PR 29-AUG-2000; 2000US-228439P.  
XX PR 29-AUG-2000; 2000US-228440P.  
XX PR 29-AUG-2000; 2000US-228441P.  
XX PR 29-AUG-2000; 2000US-228442P.  
XX PR 29-AUG-2000; 2000US-228443P.  
XX PR 29-AUG-2000; 2000US-228511P.  
XX PR 29-AUG-2000; 2000US-228512P.

```

XX WPI; 2000-452178/39.
DR N-PSDB; AAA64651, AAA64890.
XX
XX Novel polypeptides derived from Bordetella pertussis, useful for
PT treating and diagnosing Bordetella infection -
PT
XX Example 2; Pages 91-92; 165pp; English.
PS
XX
XX Bordetella pertussis possesses a type III secretion system. Type III
CC secretion systems allow bacteria to target virulence factors directly at
CC host cells. The present sequence is the BscC protein of B. pertussis.
CC The present protein is encoded by a Class I type gene and is involved in
CC the type III secretion system of B. pertussis i.e. a Bordetella
CC pathogenicity protein. The gene of the present protein is located within
CC a pathogenicity island (see AAA64890). A pathogenicity island is a
CC compact, distinct genetic unit carrying virulence genes. The present
CC protein may be used to treat or diagnose B. pertussis infection, e.g. as
CC a vaccine. Whooping cough is a disease caused by infection by B.
CC pertussis.
XX
XX
XX Sequence 600 AA;
SQ
Query Match 1.2%; Score 9; DB 21; Length 600;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 725 KVPLLGDIP 733
DB 494 KVPLLGDIP 502
|||||
|
RESULT 12
AAW53827
ID AAW53827 standard; Protein; 649 AA.
XX AAW53827;
AC
XX
XX 24-JUL-1998 (first entry)
DT
XX
XX Pseudomonas XcpQ secretion factor.
DE
XX
XX Kinase; LipQ; LipR; lipase expression regulator; DNA binding regulator;
KW sigma 54 promoter; secretion factor; lux-box binding element;
KW orfV-box binding element; regulation cascade.
XX
XX Pseudomonas alcaligenes.
OS
XX
XX WO9806836-A2.
PN
XX
XX 19-FEB-1998.
PD
XX
XX 15-AUG-1997; 97WO-US14450.
PF
XX
XX 16-AUG-1996; 96US-0699092.
PR
XX
XX (GEMV ) GENENCOOR INT INC.
PA
XX
XX Gerritse G, Quax WJ;
PI
XX
XX WPI: 1998-159528/14.
DR
XX N-PSDB; AAV23479.
DR
XX Nucleic acids encoding proteins involved in the lipase regulation
PT cascade from P. alcaligenes - useful for controlling production and
PT secretion of heterologous proteins in P. alcaligenes
XX
XX Disclosure; Page 31-33; 106pp; English.
XX
XX This sequence is the XcpQ secretion factor of Pseudomonas alcaligenes.
CC The DNA represents a nucleic acid of the invention. The nucleic acids
CC encode: (a) a kinase from a Pseudomonad that regulates the expression of
CC a lipase; (b) a DNA binding regulator from a Pseudomonad that regulates
XX
XX

```

CC the expression of a lipase; (c) a Pseudomonas alcaligenes upstream  
CC activating sequence; (d) a P. alcaligenes sigma 54 promoter that  
CC regulates the expression of a lipase; (e) a P. alcaligenes secretion  
CC factor selected from XcpP, Q, R, S, T, U, V, W, X, Y and Z, and OrfV, X,  
CC Y; (f) a P. alcaligenes lux-box binding element; and (g) a orfV-box  
CC binding element. The nucleic acids represent parts of a regulation  
CC cascade, comprising at its heart a kinase and a DNA binding regulator.  
CC These sequences can be used for the production of heterologous proteins  
CC in a host cell. The cascade also comprises secretion factors which can  
CC enhance the secretion of produced proteins.

XX Sequence 649 AA;  
SQ Query Match 1.2%; Score 9; DB 19; Length 649;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDIP 733  
DB 563 KVPLLGDIP 571  
|||||

RESULT 13  
AAY82594  
ID AAY82594 standard; Protein; 649 AA.  
XX AC AAY82594;  
XX DT 01-AUG-2000 (first entry)  
XX DE Pseudomonas alcaligenes XcpQ protein sequence SEQ ID NO:15.  
XX KW Pseudomonas alcaligenes; expression; lipase regulation cascade;  
XX KX kinase; DNA binding regulator; polymerase; promoter; secretion factor;  
XX XcpP; XcpQ; XcpR; XcpS; XcpU; XcpV; XcpW; XcpX; XcpY; XcpZ;  
XX OrfV; OrfX; OrfY; OrfZ; LipQ; LipR; upstream activating sequence;  
XX detergent; cleaning formulation.  
XX OS Pseudomonas alcaligenes.  
XX PN US6048710-A.  
XX PD 11-APR-2000.  
XX PF 15-AUG-1997; 97US-0911853.  
XX PR 16-AUG-1996; 96US-0699092.  
XX PA (GENV) GENENCOR INT INC.  
XX PI Gerritse G, Quax WJ;  
XX PS WPI; 2000-316896/27.  
XX DR N-PSDB; AAA13897.  
XX PT Expression vector for producing heterologous proteins in host cells  
XX PT which hybridizes under stringent conditions to nucleic acid -  
XX PS Example 7; Fig 3; 133pp; English.

CC The present invention describes an expression vector comprising a  
CC nucleic acid encoding a kinase and a DNA binding regulator which  
CC hybridizes under stringent conditions to a nucleic acid isolated  
CC from Pseudomonas alcaligenes. Also described are: (1) an isolated  
CC plasmid comprising the above expression vector; (2) a method of  
CC transforming a host cell comprising adding the above plasmid to host  
CC cells under appropriate conditions; (3) a transformed host cell  
CC comprising the above expression vector; and (4) a method for producing  
CC a protein comprising the steps of obtaining a host cell comprising the  
CC above expression vector and further comprising nucleic acid encoding  
CC the protein, and culturing the host cell under conditions for the  
CC expression of protein. The expression vector of the present invention

CC can be used for producing heterologous proteins in host cells,  
CC particularly, lipase in Pseudomonas. Lipases produced can be used in  
CC detergents and cleaning formulations in industrial processes. The  
CC invention provides a higher production level and efficiently express a  
CC heterologous protein. The present sequence represents XcpQ isolated from  
CC Pseudomonas alcaligenes, from the present invention.

XX Sequence 649 AA;  
SQ Query Match 1.2%; Score 9; DB 21; Length 649;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDIP 733  
DB 563 KVPLLGDIP 571  
|||||

RESULT 14  
AAB82256  
ID AAB82256 standard; Protein; 649 AA.  
XX AC AAB82256;  
XX DT 09-JUL-2001 (first entry)  
XX DE Pseudomonas alcaligenes secretion factor XcpQ.  
XX KW Lipase; XcpQ; secretion factor; protein secretion.  
XX OS Pseudomonas alcaligenes.  
XX PN US6225106-B1.  
XX PD 01-MAY-2001.  
XX PF 06-JAN-2000; 2000US-0479409.  
XX PR 15-AUG-1997; 97US-0911853.  
XX PR 16-AUG-1996; 96US-0699092.  
XX PA (GENV) GENENCOR INT INC.  
XX PI Gerritse G, Quax WJ;  
XX PS WPI; 2001-315684/33.  
XX DR N-PSDB; AAF30870.  
XX DR AAB82262; AAB82263; AAB82264; AAB82265; AAB82266; AAB82267;  
XX DR AAB82268; AAB82269.  
XX PT Novel isolated nucleic acid encoding kinase from Pseudomonas that can  
XX PT regulate expression of lipase, useful in expression systems for  
XX PT production of lipase which is useful in detergents and cleaning  
XX PT formulations -  
XX PS Example 7; Fig 3A; 133pp; English.

CC The present sequence is that of secretion factor XcpQ of  
CC Pseudomonas alcaligenes. The secretion factor is encoded by an  
CC open reading frame identified in a cosmid (see AAF30870) derived  
CC from P. alcaligenes DNA. Secretion factors aid the secretion of  
CC other proteins from a cell. A new expression system comprises  
CC components of a lipase regulation cascade including a kinase, DNA  
CC binding regulator, polymerase, a promoter, an upstream activating  
CC sequence, and secretion factors. The secretion factor is preferably  
CC a member of the Xcp protein family (see AAB82256-68) and acts in  
CC concert with other members of the Xcp family. Plasmids and  
CC transformed cells are provided, and also host cells which further  
CC comprise a nucleic acid encoding a desired protein, especially an  
CC esterase, hydrolase, lipase, isomerase, mutase, transferase, kinase  
CC or phosphatase (claimed). A hyper-producing strain can be  
CC developed.

SQ Sequence 649 AA;  
 Query Match 1.2%; Score 9; DB 22; Length 649;  
 Best Local Similarity 100.0%; Pred.No.11;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDIP 733  
 |||||  
 563 KVPLLGDIP 571

Db

RESULT 15  
 AAEL13660  
 ID AAEL13660 standard; Protein; 649 AA.  
 XX  
 AC AAEL13660;  
 XX  
 DT 26-FEB-2002 (first entry)  
 XX  
 DE Pseudomonas alcaligenes XcpQ secretion factor protein.  
 XX  
 KW DNA binding regulator; lipase regulation cascade; kinase; polymerase;  
 KW promoter; upstream activating factor; secretion factor; detergent;  
 KW cleaning formulation; industrial process; XcpQ protein.  
 XX  
 OS Pseudomonas alcaligenes.  
 XX  
 PN US63113283-B1.  
 XX  
 PD 06-NOV-2001.  
 XX  
 PF 06-JAN-2000; 2000US-0479453.  
 XX  
 PR 15-AUG-1997; 97US-0911853.  
 PR 16-AUG-1996; 96US-0699092.  
 XX  
 PA (GEMV ) GENENCOR INT INC.  
 XX  
 PI Gerritse G, Quax WJ;  
 XX  
 DR WPI; 2002-024912/03.  
 DR N-PSDB; AAD22874.  
 XX  
 PS Polynucleotide encoding DNA binding regulator, useful for regulating  
 FT expression of lipase especially in pseudomonad, and lipase useful in  
 PT detergents and other cleaning formulations and in various industrial  
 PT processes -  
 XX  
 PS Disclosure; Column 37-40; 132pp; English.  
 XX  
 CC The present invention relates to a polynucleotide encoding a DNA  
 CC binding regulator that can regulate the expression of a lipase. The  
 CC invention also relates to an expression system comprising components  
 CC of Pseudomonas alcaligenes lipase regulation cascade which includes  
 CC kinases, DNA binding regulators, polymerases, promoters, upstream  
 CC activating factors and secretion factors. DNA binding regulators  
 CC of the invention are useful for regulating the expression of a  
 CC lipase, where the lipase is useful in detergents and other cleaning  
 CC formulations as well as a number of industrial processes. Various  
 CC components of lipase regulation cascade is useful in expression  
 CC methods and systems designed for the production of heterologous  
 CC proteins. The present sequence is P. alcaligenes XcpQ, a secretion  
 CC factor protein.  
 XX  
 SQ Sequence 649 AA;

Query Match 1.2%; Score 9; DB 23; Length 649;  
 Best Local Similarity 100.0%; Pred.No.11;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDIP 733  
 |||||  
 563 KVPLLGDIP 571

Db

RESULT 16  
 AAU34768  
 ID AAU34768 standard; Protein; 654 AA.  
 XX  
 AC AAU34768;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE E. coli cellular proliferation protein #349.  
 XX  
 KW Antisense; prokaryotic cellular proliferation protein;  
 KW antibiotic; antibacterial; drug design.  
 XX  
 OS Escherichia coli.  
 XX  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlseen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 DR WPI; 2001-611495/70.  
 DR N-PSDB; AAS52627.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 PS Example 3; Seq ID No 10361; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 654 AA;

Query Match 1.2%; Score 9; DB 22; Length 654;  
 Best Local Similarity 100.0%; Pred.No.11;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDIP 733  
 |||||  
 563 KVPLLGDIP 571

Db

```

RESULT 17
ABG25804
ID   ABG25804 standard; Protein; 821 AA.
XX   AC
XX   ABG25804;
XX   DT 18-FEB-2002 (first entry)
XX   DE
XX   DE Novel human diagnostic protein #25795.
XX   KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX   KW food supplement; medical imaging; diagnostic; genetic disorder.
XX   OS Homo sapiens.
XX   PN WO200175067-A2.
XX   PD 11-OCT-2001.
XX   PF 30-MAR-2001; 2001WO-US08631.
XX   PR 31-MAR-2000; 2000US-0540217.
XX   PR 23-AUG-2000; 2000US-0649167.
XX   PA (HYSE-) HYSEQ INC.
XX   PI Drmanac RT, Liu C, Tang YT;
XX   DR WPI; 2001-639362/73.
XX   DR N-PSDB; AAS89991.
XX   PT New isolated polynucleotide and encoded polypeptides, useful in
XX   PT diagnostics, forensics, gene mapping, identification of mutations
XX   PT responsible for genetic disorders or other traits and to assess
XX   PT biodiversity.
XX   PS Claim 20; SEQ ID No 56163; 103pp; English.
XX   CC The invention relates to isolated polynucleotide (I) and
XX   CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX   CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX   CC and gene mapping, and in recombinant production of (II). The
XX   CC polynucleotides are also used in diagnostics as expressed sequence tags
XX   CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX   CC to restore normal activity of (II) or to treat disease states involving
XX   CC (III). (II) is useful for generating antibodies against it, detecting or
XX   CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX   CC a food supplement. (II) and its binding partners are useful in medical
XX   CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX   CC disorders involving aberrant protein expression or biological activity.
XX   CC The polypeptide and polynucleotide sequences have applications in
XX   CC diagnostics, forensics, gene mapping, identification of mutations
XX   CC responsible for genetic disorders or other traits to assess biodiversity
XX   CC and to produce other types of data and products dependent on DNA and
XX   CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX   CC diagnostic amino acid sequences of the invention.
XX   CC Note: The sequence data for this patent did not appear in the printed
XX   CC specification, but was obtained in electronic format directly from WIPO
XX   CC at ftp.wipo.int/pub/published_pct_sequences.
XX   SQ Sequence 821 AA;
    Query Match 1.2%; Score 9; DB 22; Length 821;
    Best Local Similarity 100.0%; Pred. No. 14;
    Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 725 KVELLGDIP 733
DB 182 KVELLGDIP 190
RESULT 18
AAW86007
ID   AAW86007 standard; Protein; 1146 AA.
XX   AC
XX   AAW86007;
XX   DT 29-MAR-1999 (first entry)
XX   DE
XX   DE Histidine kinase Cgsln1 involved in phenotypic switching.
XX   KW Cgsln1; histidine kinase; two component response regulator;
XX   KW phenotypic switching; virulence; mycosis; antimycotic; antifungal;
XX   KW fungicide; screening.
XX   OS Candida glabrata.
XX   FH Key
XX   FH Location/Qualifiers
XX   FT Domain
XX   FT /note= "membrane-spanning domain"
XX   FT 26..49
XX   FT Domain
XX   FT 321..337
XX   FT /note= "membrane-spanning domain"
XX   FT Peptide
XX   FT 493
XX   FT /note= "residue involved in signal transduction
XX   FT through transcription factor Mcmlp"
XX   FT Domain
XX   FT 518..529
XX   FT /label= H1
XX   FT Domain
XX   FT 634..638
XX   FT /label= N
XX   FT Misc-difference 752
XX   FT /note= "encoded by AAK"
XX   FT Domain
XX   FT 809..811
XX   FT /label= G1
XX   FT Domain
XX   FT 839..845
XX   FT /label= G2
XX   FT Misc-difference 841
XX   FT /note= "encoded by ACW"
XX   FT Misc-difference 842
XX   FT /note= "encoded by GGY"
XX   FT Misc-difference 844
XX   FT /note= "encoded by GGS"
XX   FT Misc-difference 846
XX   FT /note= "encoded by YTA"
XX   FT Misc-difference 1033
XX   FT /note= "encoded by GAR"
XX   FT Domain
XX   FT 1075..1080
XX   FT /label= D
XX   FT Peptide
XX   FT 1080
XX   FT /note= "residue involved in signal transduction
XX   FT through transcription factor Mcmlp"
XX   FT Peptide
XX   FT 1128
XX   FT /note= "residue involved in signal transduction
XX   FT through transcription factor Mcmlp"
XX   PN WO9855655-A1.
XX   PD 10-DEC-1998.
XX   PF 05-JUN-1998; 98WO-US11666.
XX   PR 06-JUN-1997; 97US-0048803.
XX   PA (ICWA ) UNIV IOWA RES FOUND.
XX   PI Soll DR, Srikantha T;
XX   KW WPI; 1999-059918/05.
XX   DR N-PSDB; AAV80320.
XX   PT New autophosphorylating histidine kinase gene, Cgsln1 - useful to
XX   PT screen for pharmaceutical compounds which display anti-fungal
XX   PT activity against Candida glabrata
XX   PS Disclosure; Fig 3A-B; 40pp; English.
```

XX This polypeptide comprises an autophosphorylating histidine kinase  
CC encoded by the *Candida glabrata* Cgs1n1 gene (see AA080320). Cgs1n1  
CC is a novel two component response regulator that is involved in  
CC phenotypic switching, significant because of a direct correlation  
CC between phenotypic switching and the level of virulence of the  
CC organism. A claimed method for screening candidate pharmaceutical  
CC compounds involves contacting the test substance with yeast cells  
CC containing the Cgs1n1 gene and then monitoring the effect, if any,  
CC on the level of expression of the gene. The screening method is  
CC used to identify compounds that effectively inhibit phenotypic  
CC switching, and thus pathogenicity, in *C. glabrata* for use as  
CC antimycotics.  
XX  
XX  
SQ Sequence 1146 AA;

Query Match 1.21; Score 9; DB 20; Length 1146;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 TKLTKIISG 11  
| | | | | | | |  
Db 331 TKLTKIISG 339

RESULT 19  
AAM82329  
ID AAM82329 standard; Protein; 66 AA.  
AC AAM82329;  
XX  
XX 07-NOV-2001 (first entry)  
DT  
XX  
DE Human immune/haematopoietic antigen SEQ ID NO:9922.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytosolic; gene therapy; vaccine; metastasis.  
XX  
OS Homo sapiens.  
XX  
XX WO200157182-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01354.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.

14-AUG-2000; 2000US-0225757.  
14-AUG-2000; 2000US-0225758.  
14-AUG-2000; 2000US-0225759.  
18-AUG-2000; 2000US-0226279.  
22-AUG-2000; 2000US-0226681.  
22-AUG-2000; 2000US-0226868.  
22-AUG-2000; 2000US-0227182.  
23-AUG-2000; 2000US-0227009.  
30-AUG-2000; 2000US-0228924.  
01-SEP-2000; 2000US-0229287.  
01-SEP-2000; 2000US-0229343.  
01-SEP-2000; 2000US-0229344.  
01-SEP-2000; 2000US-0229345.  
05-SEP-2000; 2000US-0229509.  
05-SEP-2000; 2000US-0229513.  
06-SEP-2000; 2000US-0230437.  
06-SEP-2000; 2000US-0230438.  
08-SEP-2000; 2000US-0231242.  
08-SEP-2000; 2000US-0231243.  
08-SEP-2000; 2000US-0231244.  
08-SEP-2000; 2000US-0231413.  
08-SEP-2000; 2000US-0231414.  
08-SEP-2000; 2000US-0232080.  
08-SEP-2000; 2000US-0232081.  
12-SEP-2000; 2000US-0231968.  
14-SEP-2000; 2000US-0232397.  
14-SEP-2000; 2000US-0232398.  
14-SEP-2000; 2000US-0232399.  
14-SEP-2000; 2000US-0232400.  
14-SEP-2000; 2000US-0232401.  
14-SEP-2000; 2000US-0233063.  
14-SEP-2000; 2000US-0233064.  
14-SEP-2000; 2000US-0233065.  
21-SEP-2000; 2000US-0234223.  
21-SEP-2000; 2000US-0234274.  
25-SEP-2000; 2000US-0234997.  
25-SEP-2000; 2000US-0234998.  
26-SEP-2000; 2000US-0234984.  
27-SEP-2000; 2000US-0235834.  
27-SEP-2000; 2000US-0235836.  
29-SEP-2000; 2000US-0236327.  
29-SEP-2000; 2000US-0236367.  
29-SEP-2000; 2000US-0236368.  
29-SEP-2000; 2000US-0236369.  
29-SEP-2000; 2000US-0236370.  
02-OCT-2000; 2000US-0236802.  
02-OCT-2000; 2000US-0237037.  
02-OCT-2000; 2000US-0237038.  
02-OCT-2000; 2000US-0237039.  
02-OCT-2000; 2000US-0237040.  
13-OCT-2000; 2000US-0239935.  
13-OCT-2000; 2000US-0239937.  
20-OCT-2000; 2000US-0240960.  
20-OCT-2000; 2000US-0241221.  
20-OCT-2000; 2000US-0241785.  
20-OCT-2000; 2000US-0241786.  
20-OCT-2000; 2000US-0241787.  
20-OCT-2000; 2000US-0241808.  
20-OCT-2000; 2000US-0241809.  
20-OCT-2000; 2000US-0241826.  
01-NOV-2000; 2000US-0244617.  
08-NOV-2000; 2000US-0246474.  
08-NOV-2000; 2000US-0246475.  
08-NOV-2000; 2000US-0246476.  
08-NOV-2000; 2000US-0246477.  
08-NOV-2000; 2000US-0246478.  
08-NOV-2000; 2000US-0246523.  
08-NOV-2000; 2000US-0246524.  
08-NOV-2000; 2000US-0246525.  
08-NOV-2000; 2000US-0246526.  
08-NOV-2000; 2000US-0246527.  
08-NOV-2000; 2000US-0246528.  
08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-483426/52.  
XX N-PSDB; AAK55110.  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
XX  
XX Claim 11; SEQ ID NO 9922; 3071pp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients' own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
XX Sequence 66 AA;  
Query Match 1.0%; Score 8; DB 22; Length 66;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 AKDKAFLO 430  
Db 18 AKDKAFLO 25  
RESULT 20  
AAB52476  
ID AAB52476 standard; protein; 111 AA.  
XX  
XX AAB52476;  
XX  
XX 23-FEB-2001 (first entry)  
XX  
XX Mycobacterium tuberculosis secreted protein #41.  
XX Mycobacterium tuberculosis secreted protein; MTSP; vaccine.  
XX Mycobacterium tuberculosis.  
XX Mycobacterium tuberculosis.  
XX WO2000066143-A1.  
XX  
XX 09-NOV-2000.  
XX  
XX 04-MAY-2000; 2000WO-US12197.  
XX  
XX 04-MAY-1999; 99US-0132479.  
XX 04-MAY-1999; 99US-0132503.  
XX  
XX (PUBL-) PUBLIC HEALTH RES INST NEW YORK.  
XX  
XX Gennaro ML, Gomez MJ;  
XX WPI; 2001-007151/01.  
XX  
XX Novel Mycobacterium tuberculosis secreted polypeptides and  
PT polynucleotides useful in diagnosis, treatment and prophylaxis of  
PT tuberculosis -  
XX  
XX Claim 11; Fig 1; 60pp; English.  
XX  
XX The present invention relates to Mycobacterium tuberculosis secreted  
CC proteins (MTSP), where the polypeptide has M. tuberculosis specific  
CC antigenic and immunogenic properties. Compositions of the invention may  
CC be useful for diagnosing Mycobacterium tuberculosis infection and as a  
CC vaccine against M. tuberculosis infection.  
XX  
XX Sequence 111 AA;  
Query Match 1.0%; Score 8; DB 22; Length 111;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 242 LAALGFAG 249  
Db 22 LAALGFAG 29  
RESULT 21  
ABP80345  
ID ABP80345 standard; protein; 118 AA.  
XX  
XX ABP80345;  
XX  
XX 07-MAR-2003 (first entry)  
XX  
XX N. gonorrhoeae amino acid sequence SEQ ID 7220.  
XX  
XX Antibacterial; infection; vaccine; gene therapy.  
XX  
XX Neisseria gonorrhoeae.  
XX  
XX WO200279243-A2.



XX 10-OCT-2002.  
 XX 12-FEB-2002; 2002WO-1B02069.  
 XX 12-FEB-2001; 2001GB-0003424.  
 XX (CHIR-) CHIRON SPA.  
 XX Fontana MR, Pizza M, Massignani V, Monaci E;  
 XX WPI; 2003-058415/05.  
 XX N-PSDB; AB241315.  
 XX New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a  
 PT medicament for treating or preventing *N. gonorrhoeae* infection -  
 PS Disclosure; Page 712; 815pp; English.  
 XX The present invention relates to proteins from *Neisseria gonorrhoeae*.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
 CC molecules of the invention.  
 XX  
 SQ Sequence 118 AA;  
 Query Match 1.0%; Score 8; DB 24; Length 118;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 AALGFAGQ 250  
 DB 99 AALGFAGQ 106  
 |||||  
 |||||

RESULT 22  
 ABG91620  
 ID ABG91620 standard; Protein; 231 AA.  
 AC ABG91620;  
 XX 18-NOV-2002 (first entry)  
 DT Purine/pyrimidine triphosphate type nucleotidyltransferase #205.  
 DE Nucleotidyltransferase; enzyme; active site engineering;  
 XX alpha-D-glucopyranosyl phosphate thymidyltransferase; Ep;  
 KW substrate specificity; nucleotide sugar;  
 KW glycosylated bioactive natural product.  
 XX  
 OS *Neisseria meningitidis*.  
 XX WO200248331-A2.  
 XX 20-JUN-2002.  
 XX 13-DEC-2001; 2001WO-US47953.  
 XX 13-DEC-2000; 2000US-254927P.  
 XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
 XX Thorson JS, Nikilov DB;  
 XX WPI; 2002-608282/65.  
 XX Nucleotidyltransferase mutated at one or more amino acids, useful in  
 PT the synthesis of nucleotide sugars -  
 XX

PS Claim 3; Page -; 182pp; English.  
 XX The invention relates to a Nucleotidyltransferase mutated at one or  
 CC more amino acids selected from V173, G147, W224, N112, G175, D111, E162,  
 CC T201, I200, E199, R195, L89, L89T, L109, Y146 or Y177 (with reference to  
 CC the *Salmonella enterica* rmlA-encoded alpha-D-glucopyranosyl phosphate  
 CC thymidyltransferase, Ep, enzyme appearing as ABG91798). The mutations  
 CC alter the substrate specificity of the enzymes. The mutants and methods  
 CC involving them are used in the synthesis of nucleotide sugars for  
 CC altering nucleotidyltransferase exhibits different substrate specificity. The  
 CC nucleotidyltransferase exhibits a non-mutated nucleotidyltransferase.  
 CC GTP, CTP, TTP, UTP and ATP than a non-mutated nucleotidyltransferase.  
 CC The mutant may also exhibit a high degree of sequence identity to  
 CC *Salmonella enterica* LT2 alpha-D-glucopyranosyl phosphate  
 CC thymidyltransferase (Ep) and can convert a wide variety of phosphates.  
 CC The mutants can be exploited in the biosynthesis of glycosylated  
 CC bioactive natural products of pharmacological use. The present  
 CC sequence is a nucleotidyltransferase exhibiting a high degree of  
 CC sequence identity to *Salmonella enterica* LT2 alpha-D-glucopyranosyl  
 CC phosphate thymidyltransferase (Ep).  
 CC Note: The present sequence is not displayed in the specification but was  
 CC obtained from Genbank.  
 XX  
 SQ Sequence 231 AA;  
 Query Match 1.0%; Score 8; DB 23; Length 231;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TAAFOQTAS 23  
 DB 116 TAAFOQTAS 123  
 |||||  
 |||||

RESULT 23  
 ABG91627  
 ID ABG91627 standard; Protein; 231 AA.  
 AC ABG91627;  
 XX 18-NOV-2002 (first entry)  
 DT Purine/pyrimidine triphosphate type nucleotidyltransferase #212.  
 DE Nucleotidyltransferase; enzyme; active site engineering;  
 KW alpha-D-glucopyranosyl phosphate thymidyltransferase; Ep;  
 KW substrate specificity; nucleotide sugar;  
 KW glycosylated bioactive natural product.  
 XX  
 OS *Neisseria meningitidis*.  
 XX WO200248331-A2.  
 XX 20-JUN-2002.  
 XX 13-DEC-2001; 2001WO-US47953.  
 XX 13-DEC-2000; 2000US-254927P.  
 XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
 XX Thorson JS, Nikilov DB;  
 XX WPI; 2002-608282/65.  
 XX Nucleotidyltransferase mutated at one or more amino acids, useful in  
 PT the synthesis of nucleotide sugars -  
 XX  
 PS Claim 3; Page -; 182pp; English.  
 XX The invention relates to a Nucleotidyltransferase mutated at one or  
 CC more amino acids selected from V173, G147, W224, N112, G175, D111, E162,  
 CC T201, I200, E199, R195, L89, L89T, L109, Y146 or Y177 (with reference to

CC the Salmonella enterica rmlA-encoded alpha-D-glucopyranosyl phosphate  
CC thymidyltransferase, Ep. enzyme appearing as ABG91798). The mutations  
CC alter the substrate specificity of the enzymes. The mutants and methods  
CC involving them are used in the synthesis of nucleotide sugars for  
CC altering nucleotidyltransferase substrate specificity. The  
CC nucleotidyltransferase exhibits different substrate specificity for  
CC GTP, CTP, TTP, and ATP than a non-mutated nucleotidyltransferase.  
CC The mutant may also exhibit a high degree of sequence identity to  
CC Salmonella enterica LT2 alpha-D-glucopyranosyl phosphate  
CC thymidyltransferase (Ep) and can convert a wide variety of phosphates.  
CC The mutants can be exploited in the biosynthesis of glycosylated  
CC bioactive natural products of pharmacological use. The present  
CC sequence is a nucleotidyltransferase exhibiting a high degree of  
CC sequence identity to Salmonella enterica LT2 alpha-D-glucopyranosyl  
CC phosphate thymidyltransferase (Ep).  
CC Note: The present sequence is not displayed in the specification but was  
CC obtained from Genbank.  
XX  
SQ Sequence 231 AA;

Query Match 1.0%; Score 8; DB 23; Length 231;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TAAQTAS 23  
DB 116 TAAQTAS 123

RESULT 24  
ABP77597  
ID ABP77597 standard; Protein; 231 AA.

AC ABP77597;  
XX  
DT 07-MAR-2003 (first entry)

DE N. gonorrhoeae amino acid sequence SEQ ID 1724.  
XX  
KW Antibacterial; infection; vaccine; gene therapy.

OS Neisseria gonorrhoeae.  
XX  
PN WO200279243-A2.  
PD 10-OCT-2002.

PF 12-FEB-2002; 2002WO-IB02069.  
PR 12-FEB-2001; 2001GB-0003424.

PA (CHIR-) CHIRON SPA.  
PI Fontana MR, Pizza M, Massignani V, Monaci E;

XX WPI; 2003-058415/05.  
DR N-PSDB; AB238567.

XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
PT medicament for treating or preventing N. gonorrhoeae infection -  
XX Disclosure; Page 312; 815pp; English.

XX The present invention relates to proteins from Neisseria gonorrhoeae.  
CC Also disclosed are the nucleic acid molecules encoding the proteins and  
CC antibodies that specifically bind to the proteins. The composition  
CC comprising the protein, nucleic acid or antibody is useful for the  
CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
CC infection, this may be in the form of a vaccine or gene therapy.  
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
CC molecules of the invention.  
XX  
SQ Sequence 231 AA;

Query Match 1.0%; Score 8; DB 24; Length 231;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TAAQTAS 23  
DB 116 TAAQTAS 123

RESULT 25  
AAU33596  
ID AAU33596 standard; Protein; 776 AA.

XX AAU33596;  
XX  
DT 14-FEB-2002 (first entry)

XX Pseudomonas aeruginosa cellular proliferation protein #40.  
XX  
KW Antisense; prokaryotic cellular proliferation protein;  
KW antibiotic; antibacterial; drug design.

OS Pseudomonas aeruginosa.  
XX  
PN WO200170955-A2.  
PD 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.  
PR 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
PA (ELIT-) ELITRA PHARM INC.

XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.  
DR N-PSDB; AAS51455.

XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX Example 3; Seq ID No 5092; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 776 AA;

Query Match 1.0%; Score 8; DB 22; Length 776;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 726 VPLLGDP 733  
 |||||  
 Db 562 VPLLGDP 569

RESULT 26  
 ABB04826  
 ID ABB04826 standard; Protein; 1024 AA.  
 AC ABB04826;  
 XX  
 DT 13-MAR-2002 (first entry)  
 XX  
 DE LDL receptor binding protein Talin SEQ ID NO:49.  
 XX  
 DE Low density lipoprotein receptor binding protein; signal transduction;  
 KW LDL receptor binding protein; LDL receptor signalling pathway.  
 KW  
 XX  
 OS Synthetic.  
 OS  
 XX WO200184159-A2.  
 XX  
 XX 08-NOV-2001.  
 XX  
 XX 24-APR-2001; 2001WO-US13214.  
 PF  
 XX  
 XX 01-MAY-2000; 2000US-0562737.  
 PR  
 XX (TEXA ) UNIV TEXAS SYSTEM.  
 PA  
 XX Herz J, Gotthardt M;  
 PI  
 XX WPI; 2002-082855/11.  
 DR  
 XX

PT Detecting stress that alters interaction of LDL receptor binding  
 PT polypeptide with LDL receptor interaction domain, comprises detecting  
 PT difference in stress-biased and unbiased interaction of peptide and  
 PT domain in a system -  
 XX  
 PS Disclosure; Page 110-112; 200pp; English.  
 XX  
 CC The present invention describes a method for detecting a stress that  
 CC alters a functional interaction of a low density lipoprotein (LDL)  
 CC receptor binding protein (I) with an LDL receptor interaction domain  
 CC (II). The method involves introducing a predetermined stress into a  
 CC system which provides a stress-biased physical interaction of (I) with  
 CC (II) where in the absence of the stress, the system provides an  
 CC unbiased interaction of (I) and (II), and detecting the stress-biased  
 CC interaction of (I) and (II), where a difference between BI and UI  
 CC indicates that the stress alters the interaction of (I) and (II). (I)  
 CC is selected from SEMCAP-1, JIP-1, PSD-95, JIP-2, Talin, OMP25, CAPON,  
 CC PIP4.5 Kinase, Na channel, brain 3, Mint1, ICAP-1 and APC subunit 10.  
 CC The method is useful for detecting a stress that alters functional  
 CC interaction of LDL receptor binding polypeptide with LDL receptor  
 CC interaction domain. The method is useful for detecting and modulating  
 CC signal transduction through LDL receptors. ABB04778 to ABB04909  
 CC represent LDL receptor binding proteins which are used in the  
 CC exemplification of the present invention.  
 XX

Query Match 1.0%; Score 8; DB 23; Length 1024;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 QAAASAKQ 201  
 |||||  
 Db 923 QAAASAKQ 930

RESULT 27  
 AAB15945  
 ID AAB15945 standard; Protein; 2393 AA.  
 XX  
 AC AAB15945;  
 XX  
 DT 05-OCT-2000 (first entry)  
 XX  
 DE E. coli proliferation associated protein sequence SEQ ID NO:302.  
 XX  
 KW Escherichia coli; E. coli; proliferation; inhibition; screening;  
 KW antimicrobial; bacterial growth; antisense therapy; antibacterial.  
 XX  
 OS Escherichia coli.  
 XX  
 PN WO2000044906-A2.  
 XX  
 PD 03-AUG-2000.  
 XX  
 PF 27-JAN-2000; 2000WO-US02200.  
 XX  
 PR 27-JAN-1999; 99US-0117405.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 DR WPI; 2000-514822/46.  
 DR N-PSDB; AAA65950.  
 XX  
 XX Novel polynucleotides and polypeptides associated with microorganism  
 PT proliferation, used to identify inhibitors of bacterial growth and  
 PT proliferation, for use in antisense therapy -  
 XX  
 PS Claim 11; Page 224-229; 316pp; English.

CC AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide  
 CC sequences derived from Escherichia coli which inhibit E. coli  
 CC proliferation. AAA65890 to AAA66055 and AAA65886 to AAA66040 represent  
 CC nucleotide and protein sequences associated with E. coli proliferation.  
 CC AAA66056 and AAA66057 represent primers used for sequencing E. coli  
 CC proliferation inhibiting nucleotide inserts in an example from the  
 CC present invention. Methods from the present invention can be used to  
 CC identify a proliferation- required gene in a microorganism, by contacting  
 CC a microorganism with a proliferation-required gene activity inhibitor  
 CC nucleic acid identified in another organism, and determining if  
 CC inhibition occurs in the second microorganism. The nucleic acid sequences  
 CC identified as being required for bacterial growth and proliferation, can  
 CC be used for antisense therapy for killing bacteria.  
 XX  
 SQ Sequence 2383 AA;

Query Match 1.0%; Score 8; DB 21; Length 2383;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 DELDVPAQ 514  
 |||||  
 Db 138 DELDVPAQ 145

RESULT 28  
 AAY62873  
 ID AAY62873 standard; Peptide; 7 AA.  
 XX  
 AC AAY62873;  
 XX  
 DT 02-MAR-2000 (first entry)  
 XX  
 DE PB-cadherin cell adhesion recognition cyclic peptide SEQ ID NO:2429.

XX Modulation; nonclassical cadherin mediated cell adhesion; CAR;  
 KW inhibition; cadherin extracellular domain; cell adhesion recognition;  
 KW OB-cadherin; cadherin-5; cadherin-6; cadherin-7; cadherin-8;  
 KW cadherin-12; cadherin-14; cadherin-15; T-cadherin; PB-cadherin;  
 KW cadherin related neuronal receptor; Li-cadherin; protocadherin;  
 KW desmoglein; desmocollin; calcium binding; cancer; tumour; obesity;  
 KW rheumatoid arthritis; multiple sclerosis; diabetes; metastasis;  
 KW neurological disease; cyclic.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 1..7  
 FT /note= "the terminal residues are condensed with each  
 FT other to form a cyclic peptide"  
 XX  
 XX  
 FN W09957149-A2.  
 XX  
 XX 11-NOV-1999.  
 XX  
 XX 05-MAY-1999; 99WO-CA00363.  
 XX  
 XX 05-MAY-1998; 98US-0073040.  
 PR 06-NOV-1998; 98US-0187859.  
 PR 20-JAN-1999; 99US-0234395.  
 PR 08-MAR-1999; 99US-0264516.  
 XX  
 XX (ADHE-) ADHEREX TECHNOLOGIES INC.  
 XX  
 XX Blaschuk OW, Gour BJ, Byers S;  
 XX  
 DR WPI; 2000-038791/03.  
 XX  
 XX New cadherin modulating agents, used for modulating nonclassical  
 PT cadherin-mediated functions for treating e.g. cancers, obesity,  
 PT rheumatoid arthritis, multiple sclerosis, diabetes or a neurological  
 PT disease  
 XX  
 XX Claim 72; Page 194; 252pp; English.  
 PS  
 CC The present invention describes cadherin modulating agents (MA)  
 CC comprising peptides which comprise a nonclassical cadherin cell adhesion  
 CC recognition (CAR) sequence. The MAs can be used for modulating  
 CC nonclassical cadherin-mediated functions. They can be used for e.g.  
 CC inhibiting adhesion of nonclassical-cadherin expressing cells in a  
 CC mammal, enhancing delivery of a drug through the skin of a mammal,  
 CC enhancing delivery of a drug to a tumour in a mammal, treating cancer in  
 CC a mammal, inhibiting metastasis of a cancer in a mammal, inhibiting  
 CC angiogenesis in a mammal, inducing apoptosis in a nonclassical cadherin-  
 CC expressing cell, preventing or treating obesity in a mammal, stimulating  
 CC blood vessel regression in a mammal, enhancing drug delivery to the  
 CC central nervous system, treating a demyelinating neurological disease,  
 CC increasing vasopermeability in a mammal, enhancing adhesion of  
 CC nonclassical cadherin-expressing cells, inhibiting synaptic stability in  
 CC a mammal, or preventing pregnancy in a mammal. They can also be used for  
 CC e.g. enhancing or directing neurite outgrowth, facilitating wound  
 CC healing or reducing scar tissue, or enhancing adhesion of foreign tissue  
 CC in a mammal. They can also be used for treating e.g. psoriasis,  
 CC arthritis, age-related macular degeneration, multiple sclerosis and  
 CC diabetes. The products can also be used for detection and diagnosis and  
 CC in bioreactors. AAY60592 to AAY64572 represent specifically claimed  
 CC peptides, and AAY64573 to AAY64643 and AAZ33183 to AAZ33186 represent  
 CC sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 7 AA;  
 Query Match 0.9%; Score 7; DB 21; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 504 KLIDELD 510

Db 1 KLIDELD 7  
 |||||  
 RESULT 29  
 AAR26158  
 ID AAR26158 standard; peptide; 10 AA.  
 XX  
 AC AAR26158;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 02-FEB-1993 (first entry)  
 XX  
 DE Recombinant signal peptide #4.  
 XX  
 KW Translocation; 85-C; Mycobacterium tuberculosis; immunise; vaccine.  
 XX  
 OS Synthetic.  
 XX  
 FN EP499003-A1.  
 XX  
 PD 19-AUG-1992.  
 XX  
 PF 14-FEB-1991; 91EP-0400388.  
 XX  
 PR 14-FEB-1991; 91EP-0400388.  
 XX  
 PA (INNO-) INNOGENETICS NV SA.  
 XX  
 XX Content J, De Bruyn J, De Wit L;  
 XX  
 XX WPI; 1992-277793/34.  
 DR  
 XX Recombinant peptide(s) and their nucleic acids - for diagnosing  
 PT tuberculosis and as a vaccine against tuberculosis  
 XX  
 XX Claim 1; Page 20; 48pp; English.  
 XX  
 CC The sequences given in AAR26155-59 are peptides which were used in the  
 CC scope of the invention as signal peptides. Their role is to  
 CC initiate the translocation of a protein from the site of synthesis,  
 CC but they are excised during translocation. The peptides which are  
 CC translated by these signal peptides are a new group of proteins  
 CC which can be used for the detection and control of tuberculosis.  
 CC They correspond to the 85-C antigen containing region of  
 CC Mycobacterium tuberculosis and can be used in the production of  
 CC vaccines for immunisation against tuberculosis.  
 CC (Updated on 25-MAR-2003 to correct PN field.)  
 XX  
 XX Sequence 10 AA;  
 SQ  
 Query Match 0.9%; Score 7; DB 13; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 141 AAPAAPA 147  
 |||||  
 DB 3 AAPAAPA 9  
 RESULT 30  
 AAG96038  
 ID AAG96038 standard; Peptide; 10 AA.  
 XX  
 AC AAG96038;  
 XX  
 DT 18-SEP-2001 (first entry)  
 XX  
 DE Human complementary peptide, SEQ ID NO: 2232.  
 XX  
 KW Human; complementary peptide; ligand; drug discovery; drug design.  
 XX  
 OS Homo sapiens.

```

XX PN WO200142277-A2.
XX PD 14-JUN-2001.
XX PF 13-DEC-2000; 2000WO-GB04776.
XX PR 13-DEC-1999; 99GB-0029464.
XX PA (PROT-) PROTEOM LTD.
XX PI Roberts GW, Heal JR;
XX PT WPI; 2001-408419/43.
XX DR
XX PT A set of peptide ligands consisting of specific complementary peptides
XX PT to proteins encoded by genes of the human genome, useful in an assay
XX PT for screening and identifying of one or more novel peptides which are
XX PT drug candidates or pro-drugs.
XX PS Example 4; Page 362; 646pp; English.
XX CC The invention relates to a set of complementary peptide ligands
XX CC generated from the human genome. The complementary peptides
XX CC interact with their relevant target proteins encoded in the human
XX CC genome. They can be used as reagents in drug discovery and as lead
XX CC ligands to facilitate drug design and development. The present
XX CC sequence is a complementary peptide provided in the specification.
XX CC Sequence 10 AA;
XX QY 141 AAPAAPA 147
XX DB 3 AAPAAPA 9

Query Match 0.9%; Score 7; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 141 AAPAAPA 147
XX DB 3 AAPAAPA 9

RESULT 31
AAG96040
ID AAG96040 standard; Peptide; 10 AA.
XX AC AAG96040;
XX DT 18-SEP-2001 (first entry)
XX DE Human complementary peptide, SEQ ID NO: 2234.
XX KW Human; complementary peptide; ligand; drug discovery; drug design.
XX OS Homo sapiens.
XX PN WO200142277-A2.
XX PD 14-JUN-2001.
XX PF 13-DEC-2000; 2000WO-GB04776.
XX PR 13-DEC-1999; 99GB-0029464.
XX PA (PROT-) PROTEOM LTD.
XX PI Roberts GW, Heal JR;
XX PT WPI; 2001-408419/43.
XX DR
XX PT A set of peptide ligands consisting of specific complementary peptides
XX PT to proteins encoded by genes of the human genome, useful in an assay
XX PT for screening and identifying of one or more novel peptides which are
XX PT drug candidates or pro-drugs.
XX PS Example 4; Page 362; 646pp; English.
XX CC The invention relates to a set of complementary peptide ligands
XX CC generated from the human genome. The complementary peptides
XX CC interact with their relevant target proteins encoded in the human
XX CC genome. They can be used as reagents in drug discovery and as lead
XX CC ligands to facilitate drug design and development. The present
XX CC sequence is a complementary peptide provided in the specification.
XX CC Sequence 10 AA;
XX QY 141 AAPAAPA 147
XX DB 3 AAPAAPA 9

Query Match 0.9%; Score 7; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 141 AAPAAPA 147
XX DB 3 AAPAAPA 9

RESULT 32
AAG96222
ID AAG96222 standard; Peptide; 10 AA.
XX AC AAG96222;
XX DT 18-SEP-2001 (first entry)
XX DE Human complementary peptide, SEQ ID NO: 2416.
XX KW Human; complementary peptide; ligand; drug discovery; drug design.
XX OS Homo sapiens.
XX PN WO200142277-A2.
XX PD 14-JUN-2001.
XX PF 13-DEC-2000; 2000WO-GB04776.
XX PR 13-DEC-1999; 99GB-0029464.
XX PA (PROT-) PROTEOM LTD.
XX PI Roberts GW, Heal JR;
XX PT WPI; 2001-408419/43.
XX DR
XX PT A set of peptide ligands consisting of specific complementary peptides
XX PT to proteins encoded by genes of the human genome, useful in an assay
XX PT for screening and identifying of one or more novel peptides which are
XX PT drug candidates or pro-drugs.
XX PS Example 4; Page 388; 646pp; English.
XX CC The invention relates to a set of complementary peptide ligands
XX CC generated from the human genome. The complementary peptides
XX CC interact with their relevant target proteins encoded in the human
XX CC genome. They can be used as reagents in drug discovery and as lead
XX CC ligands to facilitate drug design and development. The present
XX CC sequence is a complementary peptide provided in the specification.
XX CC Sequence 10 AA;
XX QY 141 AAPAAPA 147
XX DB 3 AAPAAPA 9

Query Match 0.9%; Score 7; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 141 AAPAAPA 147
XX DB 3 AAPAAPA 9

A set of peptide ligands consisting of specific complementary peptides
to proteins encoded by genes of the human genome, useful in an assay
for screening and identifying of one or more novel peptides which are
drug candidates or pro-drugs.

```

```

PS Example 4; Page 362; 646pp; English.
XX CC The invention relates to a set of complementary peptide ligands
XX CC generated from the human genome. The complementary peptides
XX CC interact with their relevant target proteins encoded in the human
XX CC genome. They can be used as reagents in drug discovery and as lead
XX CC ligands to facilitate drug design and development. The present
XX CC sequence is a complementary peptide provided in the specification.
XX CC Sequence 10 AA;
XX QY 141 AAPAAPA 147
XX DB 4 AAPAAPA 10

Query Match 0.9%; Score 7; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 141 AAPAAPA 147
XX DB 4 AAPAAPA 10

RESULT 32
AAG96222
ID AAG96222 standard; Peptide; 10 AA.
XX AC AAG96222;
XX DT 18-SEP-2001 (first entry)
XX DE Human complementary peptide, SEQ ID NO: 2416.
XX KW Human; complementary peptide; ligand; drug discovery; drug design.
XX OS Homo sapiens.
XX PN WO200142277-A2.
XX PD 14-JUN-2001.
XX PF 13-DEC-2000; 2000WO-GB04776.
XX PR 13-DEC-1999; 99GB-0029464.
XX PA (PROT-) PROTEOM LTD.
XX PI Roberts GW, Heal JR;
XX PT WPI; 2001-408419/43.
XX DR
XX PT A set of peptide ligands consisting of specific complementary peptides
XX PT to proteins encoded by genes of the human genome, useful in an assay
XX PT for screening and identifying of one or more novel peptides which are
XX PT drug candidates or pro-drugs.
XX PS Example 4; Page 388; 646pp; English.
XX CC The invention relates to a set of complementary peptide ligands
XX CC generated from the human genome. The complementary peptides
XX CC interact with their relevant target proteins encoded in the human
XX CC genome. They can be used as reagents in drug discovery and as lead
XX CC ligands to facilitate drug design and development. The present
XX CC sequence is a complementary peptide provided in the specification.
XX CC Sequence 10 AA;
XX QY 141 AAPAAPA 147
XX DB 3 AAPAAPA 9

Query Match 0.9%; Score 7; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 141 AAPAAPA 147
XX DB 3 AAPAAPA 9

A set of peptide ligands consisting of specific complementary peptides
to proteins encoded by genes of the human genome, useful in an assay
for screening and identifying of one or more novel peptides which are
drug candidates or pro-drugs.

```

RESULT 33  
 AAG96224  
 ID AAG96224 standard; Peptide; 10 AA.  
 XX AC AAG96224;  
 XX DT 18-SEP-2001 (first entry)  
 XX DE Human complementary peptide, SEQ ID NO: 2418.  
 XX KW Human; complementary peptide; ligand; drug discovery; drug design.  
 XX OS Homo sapiens.  
 XX PN WO200142277-A2.  
 XX PD 14-JUN-2001.  
 XX PF 13-DEC-2000; 2000WO-GB04776.  
 XX PR 13-DEC-1999; 99GB-0029464.  
 XX PS (PROT-) PROTEOM LTD.  
 XX PI Roberts GW, Heal JR;  
 XX DR WPI; 2001-408419/43.  
 XX PT A set of peptide ligands consisting of specific complementary peptides  
 PT to proteins encoded by genes of the human genome, useful in an assay  
 PT for screening and identifying of one or more novel peptides which are  
 PT drug candidates or pro-drugs -  
 XX PS Example 4; Page 388; 646pp; English.  
 XX CC The invention relates to a set of complementary peptide ligands  
 CC generated from the human genome. The complementary peptides  
 CC interact with their relevant target proteins encoded in the human  
 CC genome. They can be used as reagents in drug discovery and as lead  
 CC ligands to facilitate drug design and development. The present  
 CC sequence is a complementary peptide provided in the specification.  
 XX SQ Sequence 10 AA;  
 Query Match 0.9%; Score 7; DB 22; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 27; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0;  
 QY 141 AAPAAPA 147  
 |||||  
 Db 4 AAPAAPA 10  
 RESULT 34  
 ABP82732  
 ID ABP82732 standard; Peptide; 19 AA.  
 XX AC ABP82732;  
 XX DT 04-MAR-2003 (first entry)  
 XX DE G protein-coupled receptor (GPCR) antigenic peptide SEQ ID NO:1405.  
 XX KW G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;  
 KW G protein-coupled receptor modulator; antibody; immune-related disease;  
 KW growth-related disease; cell regeneration-related disease; AIDS; cancer;  
 KW immunological-related cell proliferative disease; autoimmune disease;  
 KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;  
 KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;  
 KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;  
 KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;  
 KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;  
 KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;  
 KW ulcer.  
 XX OS Homo sapiens.  
 XX PN WO200261087-A2.  
 XX PD 08-AUG-2002.  
 XX PF 19-DEC-2001; 2001WO-US50107.  
 XX PR 19-DEC-2000; 2000US-257144P.  
 XX PS (LIFE-) LIFESPAN BIOSCIENCES INC.  
 XX PI Burmer GC, Roush CL, Brown JP;  
 XX DR WPI; 2003-046718/04.  
 XX PT New isolated antigenic peptides e.g., for G protein-coupled receptors  
 PT (GPCR), useful for diagnosing and designing drugs for treating  
 PT conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease,  
 PT cancer or autoimmune diseases -  
 XX PS Claim 1; Fig 2; 523pp; English.  
 XX CC The present invention describes antigenic peptides (I) comprising:  
 CC (a) any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino  
 CC acids. Also described: (1) an assay for the detection of a particular  
 CC G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;  
 CC and (2) an isolated antibody having high specificity and high affinity  
 CC or avidity for a particular GPCR. (I) can be used as GPCR modulators and  
 CC in gene therapy. The antigenic peptides for GPCRs are useful in detecting  
 CC an antibody against a particular GPCR, and in the production of specific  
 CC antibodies. The peptides and antibodies are also useful for detecting the  
 CC presence or absence of corresponding GPCRs. The antigenic peptides for  
 CC GPCRs and antibodies are useful for diagnosing and designing drugs for  
 CC treating immune-related diseases, growth-related diseases, cell  
 CC regeneration-related disease, immunological-related cell proliferative  
 CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,  
 CC atherosclerosis, bacterial, fungal, protozoan or viral infections,  
 CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute  
 CC inflammation, allergies, Crohn's disease, diabetes, graft versus host  
 CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,  
 CC anxiety, depression, schizophrenia, dementia, mental retardation, memory  
 CC loss, epilepsy, asthma, tuberculosis, rheumatoid arthritis, trauma, ulcers, or  
 CC hypotension, renal disorders, rheumatoid arthritis. The antibodies may be  
 CC any other disorder in which GPCRs are involved. The antibodies may be  
 CC used in immunoassays and immunodiagnosis. ABZ42523 to ABZ42869 encode  
 CC GPCR proteins given in ABP81675 to ABP82018, which are used in the  
 CC exemplification of the present invention.  
 XX SQ Sequence 19 AA;  
 Query Match 0.9%; Score 7; DB 24; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 330 NLESGGV 336  
 |||||  
 Db 7 NLESGGV 13  
 RESULT 35  
 AAR26166  
 ID AAR26166 standard; peptide; 20 AA.  
 XX AC AAR26166;  
 XX DT 25-MAR-2003 (updated)  
 XX DT 02-FEB-1993 (first entry)  
 XX KW Tuberculosis antibody production peptide #7.  
 XX DE

KW 85-C, Mycobacterium tuberculosis; immunise; antibody; vaccine;  
 KW carrier molecule; antiserum.  
 XX Synthetic.  
 OS  
 XX  
 XX  
 PN EP499003-A1.  
 XX  
 PD 19-AUG-1992.  
 XX  
 PF 14-FEB-1991; 91EP-0400388.  
 XX  
 XX 14-FEB-1991; 91EP-0400388.  
 PR  
 XX (INNO-) INNOGENETICS NV SA.  
 PA  
 XX Content J, De Bruyn J, De Wit L;  
 XX  
 PI  
 XX  
 XX  
 DR WPI; 1992-277793/34.  
 XX  
 XX Recombinant peptide(s) and their nucleic acids - for diagnosing  
 PT tuberculosis and as a vaccine against tuberculosis  
 PT  
 XX  
 XX Disclosure; Page 11; 48pp; English.  
 PS  
 XX  
 XX The sequences given in AAR26160-66 are peptides which were used in the  
 CC scope of the invention to raise antibodies against tuberculosis.  
 CC They correspond to regions of the 85-C antigen containing region of  
 CC Mycobacterium tuberculosis and can be used in the production of  
 CC vaccines for immunisation against tuberculosis. The peptides may  
 CC be used to raise antisera and in this case would be synthesised with an  
 CC additional cysteine residue, pref. attached to the amino terminal.  
 CC This facilitates coupling of the peptide to a carrier molecule which  
 CC is necessary to render the peptide immunogenic.  
 CC (Updated on 25-MAR-2003 to correct PN field.)  
 CC  
 XX  
 XX  
 SQ Sequence 20 AA;

Query Match 0.9%; Score 7; DB 13; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 50;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 141 AAPAAPA 147  
 |||||  
 Db 13 AAPAAPA 19

RESULT 36  
 AAB74443  
 ID AAB74443 standard; peptide; 29 AA.  
 AC AAB74443;  
 XX  
 XX 29-MAY-2001 (first entry)  
 DT  
 XX Herpes simplex virus 2 glycoprotein B fragment #1.  
 DE  
 XX HSV-1; HSV-2; glycoprotein B; GB; transmembrane envelope glycoprotein;  
 KW antigenic epitope; diagnosis; vaccine.  
 KW  
 XX Herpes simplex virus type 2.  
 OS  
 XX  
 XX US6197497-B1.  
 PN  
 XX  
 XX 06-MAR-2001.  
 PD  
 XX  
 XX 19-APR-1996; 96US-0632537.  
 PF  
 XX 21-APR-1995; 95US-0426604.  
 PR  
 XX (UYNE-) UNIV NEW MEXICO STATE.  
 PA  
 XX Goade DE, Bell R, Jenison S;  
 PI  
 XX

DR WPI; 2001-256360/26.  
 XX Continuous, isolated, antigenic polypeptide segment of herpes simplex  
 PT virus (HSV) glycoprotein B1 or B2, useful in serodiagnostic  
 PT immunoassays for distinguishing HSV-1 infection from HSV-2 in a human -  
 XX  
 XX Examples; Fig 2; 23pp; English.  
 PS  
 XX The present invention provides antigenic peptides from herpes simplex  
 CC virus type 1 (HSV1) and 2 (HSV2) glycoprotein B (GB) which can be used in  
 CC the diagnosis of HSV infection, and identification of subtype, and in  
 CC vaccines to protect against HSV. The present sequence is a fragment of  
 CC the HSV-2 GB protein.  
 CC  
 XX  
 XX Sequence 29 AA;

Query Match 0.9%; Score 7; DB 22; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 141 AAPAAPA 147  
 |||||  
 Db 2 AAPAAPA 8

RESULT 37  
 AAR77779  
 ID AAR77779 standard; Peptide; 38 AA.  
 XX  
 AC AAR77779;  
 XX  
 XX 24-JAN-1996 (first entry)  
 DT  
 XX Goldfish derived osteoporosis treatment peptide.  
 DE  
 XX Goldfish; osteoporosis; osteoblasts; pro-osteoblasts; activation;  
 KW differentiation; treatment.  
 KW  
 XX Caurassius auratus.  
 OS  
 XX JP07118296-A.  
 PN  
 XX 09-MAY-1995.  
 PD  
 XX 22-OCT-1993; 93JP-0265163.  
 PF  
 XX 22-OCT-1993; 93JP-0265163.  
 PR  
 XX (SUMQ ) SUMITOMO METAL IND LTD.  
 PA  
 XX WPI: 1995-203835/27.  
 DR N-PSDB; A9Q94405.  
 DR  
 XX Peptide for osteoporosis treatment - has (pro-) osteoblast  
 PT differentiation and activation effect, isolated from sardine or  
 PT goldfish  
 PT  
 XX Claim 1; Page 6; 9pp; Japanese.  
 PS  
 XX AAQ94405 encodes AAR77779 a goldfish derived osteoporosis treatment  
 CC peptide. The peptide has a (pro-) osteoblast differentiation and  
 CC activation effect, and a pref. dosage of 0.1-5mg/kg/day.  
 CC  
 XX  
 XX Sequence 38 AA;

Query Match 0.9%; Score 7; DB 16; Length 38;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 133 APAREAV 139  
 |||||  
 Db 1 APAREAV 7

RESULT 38  
AA58555  
ID AAY58555 standard; peptide; 42 AA.  
XX  
AC AAY58555;  
XX  
DT 10-APR-2000 (first entry)  
XX  
DE Rat/mouse somatostatin receptor SST1 antigenic region, SEQ ID NO:23.  
XX  
DE Somatostatin receptor; SSR; receptor subtype; peptide antigen;  
KW immunogen; antibody; detection.  
XX  
OS Rattus sp.  
XX  
OS Mus sp.  
XX  
PN US5998154-A.  
XX  
XX 07-DEC-1999.  
XX  
XX 07-JUL-1995; 95US-0499676.  
XX  
XX 07-JUL-1995; 95US-0499676.  
XX  
XX (TEXA ) UNIV TEXAS SYSTEM.  
XX  
XX Schonbrunn A;  
XX  
XX WPI; 2000-104607/09.  
XX  
XX Producing somatostatin receptor subtype specific antibodies for  
PT therapeutic and diagnostic use -  
XX  
XX Disclosure; Fig 2; 31pp; English.  
XX  
XX Sequences AAY58533-Y58572 represent antigenic peptides derived from  
CC somatostatin receptor (SSR) subtypes of rat, mouse and human. The  
CC invention relates to peptide antigens derived from a specific antigenic  
CC region of a somatostatin receptor, to SSR subtype-specific antibodies  
CC induced by the peptide antigens, and to immunological methods using the  
CC SSR-subtype specific antibodies for determining the presence and  
CC distribution of SSR subtypes in a tissue sample, and for the isolation  
CC and purification of somatostatin receptor proteins. The antigenic region  
CC of SSR is located in the C-terminus, which extends intracellularly from  
CC transmembrane domain 7. The peptide antigens that are derived from this  
CC region are capable of inducing high affinity, high specificity anti-  
CC somatostatin antibodies which can distinguish between SSR subtypes. The  
CC methods of the invention may be used for characterisation of the SSR  
CC subtypes that are expressed by the cells of a tissue sample e.g., a  
CC tumour sample, via immunological methods such as immunoblot,  
CC immunocytochemical or immunoprecipitation analysis. Identification  
CC of SSR subtypes is important in the development of subtype-specific  
CC somatostatin analogues and for their diagnostic and therapeutic use.  
XX  
SQ Sequence 42 AA;  
XX  
Query Match 0.9%; Score 7; DB 21; Length 42;  
Best Local Similarity 100.0%; Pred.No. 96;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 330 NLESGGV 336  
| | | | |  
Db 23 NLESGGV 29  
XX  
RESULT 39  
AA58556  
ID AAY58556 standard; peptide; 42 AA.  
XX  
AC AAY58556;  
XX  
DT 10-APR-2000 (first entry)

XX  
DE Human somatostatin receptor SST1 antigenic region, SEQ ID NO:24.  
XX  
KW Somatostatin receptor; SSR; receptor subtype; peptide antigen;  
KW immunogen; antibody; detection.  
XX  
OS Homo sapiens.  
XX  
PN US5998154-A.  
XX  
PD 07-DEC-1999.  
XX  
PF 07-JUL-1995; 95US-0499676.  
XX  
XX 07-JUL-1995; 95US-0499676.  
XX  
XX (TEXA ) UNIV TEXAS SYSTEM.  
XX  
XX Schonbrunn A;  
XX  
XX WPI; 2000-104607/09.  
XX  
XX Producing somatostatin receptor subtype specific antibodies for  
PT therapeutic and diagnostic use -  
XX  
XX Disclosure; Fig 2; 31pp; English.  
XX  
XX Sequences AAY58533-Y58572 represent antigenic peptides derived from  
CC somatostatin receptor (SSR) subtypes of rat, mouse and human. The  
CC invention relates to peptide antigens derived from a specific antigenic  
CC region of a somatostatin receptor, to SSR subtype-specific antibodies  
CC induced by the peptide antigens, and to immunological methods using the  
CC SSR-subtype specific antibodies for determining the presence and  
CC distribution of SSR subtypes in a tissue sample, and for the isolation  
CC and purification of somatostatin receptor proteins. The antigenic region  
CC of SSR is located in the C-terminus, which extends intracellularly from  
CC transmembrane domain 7. The peptide antigens that are derived from this  
CC region are capable of inducing high affinity, high specificity anti-  
CC somatostatin antibodies which can distinguish between SSR subtypes. The  
CC methods of the invention may be used for characterisation of the SSR  
CC subtypes that are expressed by the cells of a tissue sample e.g., a  
CC tumour sample, via immunological methods such as immunoblot,  
CC immunocytochemical or immunoprecipitation analysis. Identification  
CC of SSR subtypes is important in the development of subtype-specific  
CC somatostatin analogues and for their diagnostic and therapeutic use.  
XX  
SQ Sequence 42 AA;  
XX  
Query Match 0.9%; Score 7; DB 21; Length 42;  
Best Local Similarity 100.0%; Pred.No. 96;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 330 NLESGGV 336  
| | | | |  
Db 23 NLESGGV 29  
XX  
RESULT 40  
AAB16556  
ID AAB16556 standard; Protein; 43 AA.  
XX  
AC AAB16556;  
XX  
XX 27-OCT-2000 (first entry)  
XX  
XX Bacteriophage 192 protein sequence 182ORF055.  
XX  
XX Bacteriophage; antimicrobial; genome; identification; antibacterial;  
KW bacterial growth inhibition; bacterial infection.  
XX  
XX Bacteriophage 182.  
XX  
XX WO200032825-A2.  
PN



```

XX 08-JUN-2000.
XX
XX 03-DEC-1999; 99WO-1B02040.
XX
XX 03-DEC-1998; 98US-0110992.
XX 03-JUN-1999; 99US-0326144.
XX 28-SEP-1999; 99US-0407804.
XX 30-SEP-1999; 99US-0157218.
XX 01-DEC-1999; 99US-0168777.
XX 02-DEC-1999; 99US-0454252.
XX
XX (PHAG-) PHAGETECH INC.
XX
XX Pelletier J, Gros P, Dubow M;
XX
XX WPI; 2000-412361/35.
XX N-PSDB; AAA69142.
XX
XX Identifying a bacteriophage coding region for treating bacterial
XX infections comprises identifying a nucleic acid encoding a product that
XX inhibits bacteria when a bacteriophage infects a bacterium
XX
XX Example 13; Page 326; 456pp; English.
XX
XX The present invention describes a method for identifying a bacteriophage
XX coding region encoding a product active on an essential bacterial
XX target. The method comprises identifying a nucleic acid sequence encoding
XX a gene product that provides a bacteria-inhibiting function when an
XX uncharacterised bacteriophage infects a pathogenic bacterium. The
XX compound active on a target of a bacteriophage inhibitor protein in a
XX bacteria is used to treat or prevent a bacterial infection in an animal.
XX AAA68243 to AAA69442 and AAB16523 to AAB16954 represent bacteriophage
XX nucleotide and protein sequences which are used in the exemplification of
XX the present invention.
XX
XX Sequence 43 AA;
SQ
Query Match 0.9%; Score 7; DB 21; Length 43;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 TKLTKII 9
Db 13 TKLTKII 19

```

Search completed: December 9, 2003, 10:34:08  
 Job time : 65 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 10:34:14 ; Search time 21 seconds  
(without alignments)  
1549.383 Million cell updates/sec

Title: US-09-701-271A-2  
Perfect score: 769  
Sequence: 1 MNTXKTIISGLFVATAAFQ.....ELLIFITPRIMGTAGNSLRY 769

Scoring table: OLIGO Gapop 60.0 , Gapext 60.0  
328717 seqs, 42310858 residues

Searched: 0  
Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description          |
|------------|-------|-------------|--------|----|----------------------|
| 1          | 11    | 1.4         | 739    | 4  | US-09-328-352-6048   |
| 2          | 11    | 1.4         | 751    | 4  | US-09-252-991A-33073 |
| 3          | 9     | 1.2         | 649    | 3  | US-08-911-853-15     |
| 4          | 9     | 1.2         | 649    | 3  | US-09-479-409-15     |
| 5          | 9     | 1.2         | 649    | 4  | US-09-479-453-15     |
| 6          | 9     | 1.2         | 733    | 4  | US-09-252-991A-28934 |
| 7          | 8     | 1.0         | 828    | 4  | US-09-252-991A-30225 |
| 8          | 8     | 1.0         | 1024   | 4  | US-09-562-737-49     |
| 9          | 7     | 0.9         | 7      | 4  | US-09-187-859-2429   |
| 10         | 7     | 0.9         | 7      | 4  | US-09-839-542B-2429  |
| 11         | 7     | 0.9         | 10     | 2  | US-08-716-317-32     |
| 12         | 7     | 0.9         | 10     | 2  | US-08-716-317-33     |
| 13         | 7     | 0.9         | 10     | 2  | US-08-107-676-7      |
| 14         | 7     | 0.9         | 20     | 2  | US-08-107-676-22     |
| 15         | 7     | 0.9         | 33     | 4  | US-09-194-296-36     |
| 16         | 7     | 0.9         | 42     | 2  | US-08-499-676A-23    |
| 17         | 7     | 0.9         | 42     | 2  | US-08-499-676A-24    |
| 18         | 7     | 0.9         | 50     | 3  | US-08-905-223-472    |
| 19         | 7     | 0.9         | 70     | 2  | US-08-499-676A-1     |
| 20         | 7     | 0.9         | 70     | 2  | US-08-499-676A-2     |
| 21         | 7     | 0.9         | 70     | 2  | US-08-499-676A-3     |
| 22         | 7     | 0.9         | 79     | 4  | US-09-352-991A-23964 |
| 23         | 7     | 0.9         | 85     | 4  | US-09-252-991A-21927 |
| 24         | 7     | 0.9         | 88     | 4  | US-09-252-991A-21928 |
| 25         | 7     | 0.9         | 112    | 1  | US-07-754-918A-12    |
| 26         | 7     | 0.9         | 124    | 4  | US-09-252-991A-20183 |
| 27         | 7     | 0.9         | 130    | 4  | US-08-485-116A-7     |
| 28         | 7     | 0.9         | 139    | 4  | US-09-252-991A-28069 |
| 29         | 7     | 0.9         | 145    | 4  | US-09-252-991A-31284 |
| 30         | 7     | 0.9         | 154    | 4  | US-09-252-991A-25515 |
| 31         | 7     | 0.9         | 166    | 2  | US-08-775-978-1      |
| 32         | 7     | 0.9         | 179    | 4  | US-09-252-991A-23675 |
| 33         | 7     | 0.9         | 188    | 4  | US-09-489-847-314    |
| 34         | 7     | 0.9         | 197    | 4  | US-09-252-991A-30359 |
| 35         | 7     | 0.9         | 216    | 4  | US-09-107-532A-4916  |
| 36         | 7     | 0.9         | 218    | 3  | US-08-837-058-4      |
| 37         | 7     | 0.9         | 218    | 3  | US-08-841-889-4      |
| 38         | 7     | 0.9         | 218    | 4  | US-09-417-264-4      |
| 39         | 7     | 0.9         | 221    | 4  | US-09-252-991A-18193 |
| 40         | 7     | 0.9         | 222    | 4  | US-09-252-991A-27628 |
| 41         | 7     | 0.9         | 223    | 4  | US-09-171-461-11     |
| 42         | 7     | 0.9         | 228    | 4  | US-09-107-532A-3896  |
| 43         | 7     | 0.9         | 230    | 4  | US-09-107-532A-3897  |
| 44         | 7     | 0.9         | 245    | 3  | US-09-140-804-4      |
| 45         | 7     | 0.9         | 245    | 4  | US-09-686-838B-4     |
| 46         | 7     | 0.9         | 245    | 4  | US-09-911-176B-49    |
| 47         | 7     | 0.9         | 245    | 4  | US-09-619-740-52     |
| 48         | 7     | 0.9         | 267    | 1  | US-08-416-336-2      |
| 49         | 7     | 0.9         | 322    | 1  | US-08-118-270-75     |
| 50         | 7     | 0.9         | 322    | 5  | PCT-US93-08528-75    |
| 51         | 7     | 0.9         | 325    | 4  | US-09-252-991A-23890 |
| 52         | 7     | 0.9         | 329    | 4  | US-09-107-532A-3759  |
| 53         | 7     | 0.9         | 340    | 2  | US-08-107-676-3      |
| 54         | 7     | 0.9         | 340    | 2  | US-08-107-676-31     |
| 55         | 7     | 0.9         | 340    | 2  | US-08-997-080-37     |
| 56         | 7     | 0.9         | 340    | 2  | US-08-997-362-37     |
| 57         | 7     | 0.9         | 340    | 3  | US-08-873-970-37     |
|            |       |             |        |    |                      |

|     |   |     |      |   |                      |                    |     |   |     |    |   |                    |                    |
|-----|---|-----|------|---|----------------------|--------------------|-----|---|-----|----|---|--------------------|--------------------|
| 101 | 7 | 0.9 | 637  | 4 | US-09-390-224-2      | Sequence 2, Appli  | 174 | 6 | 0.8 | 12 | 1 | US-08-320-373-21   | Sequence 21, Appl  |
| 102 | 7 | 0.9 | 637  | 4 | US-09-390-224-2      | Sequence 4, Appli  | 175 | 6 | 0.8 | 12 | 1 | US-08-320-373-26   | Sequence 26, Appl  |
| 103 | 7 | 0.9 | 643  | 3 | US-09-233-191-33     | Sequence 33, Appl  | 176 | 6 | 0.8 | 12 | 3 | US-09-171-654-43   | Sequence 43, Appl  |
| 104 | 7 | 0.9 | 643  | 3 | US-09-233-200-83     | Sequence 83, Appl  | 177 | 6 | 0.8 | 13 | 1 | US-08-320-373-15   | Sequence 15, Appl  |
| 105 | 7 | 0.9 | 643  | 4 | US-09-233-197-83     | Sequence 83, Appl  | 178 | 6 | 0.8 | 13 | 1 | US-08-320-373-20   | Sequence 20, Appl  |
| 106 | 7 | 0.9 | 643  | 4 | US-09-233-201-83     | Sequence 83, Appl  | 179 | 6 | 0.8 | 13 | 1 | US-08-320-373-24   | Sequence 24, Appl  |
| 107 | 7 | 0.9 | 656  | 4 | US-09-253-991A-18768 | Sequence 18768, A  | 180 | 6 | 0.8 | 13 | 1 | US-08-261-206A-25  | Sequence 25, Appl  |
| 108 | 7 | 0.9 | 668  | 1 | US-08-468-036-5      | Sequence 5, Appli  | 181 | 6 | 0.8 | 13 | 4 | US-08-766-596A-40  | Sequence 40, Appl  |
| 109 | 7 | 0.9 | 668  | 1 | US-08-376-843-5      | Sequence 5, Appli  | 182 | 6 | 0.8 | 13 | 4 | US-08-320-373-16   | Sequence 16, Appl  |
| 110 | 7 | 0.9 | 677  | 4 | US-09-253-991A-22442 | Sequence 22442, A  | 183 | 6 | 0.8 | 14 | 1 | US-08-320-373-19   | Sequence 19, Appl  |
| 111 | 7 | 0.9 | 741  | 1 | US-08-277-231A-4     | Sequence 4, Appli  | 184 | 6 | 0.8 | 14 | 2 | US-08-557-309B-62  | Sequence 62, Appl  |
| 112 | 7 | 0.9 | 741  | 2 | US-08-473-750-7      | Sequence 7, Appli  | 185 | 6 | 0.8 | 14 | 2 | US-08-557-309B-63  | Sequence 63, Appl  |
| 113 | 7 | 0.9 | 741  | 2 | US-08-473-326-7      | Sequence 7, Appli  | 186 | 6 | 0.8 | 14 | 2 | US-08-929-414-14   | Sequence 14, Appl  |
| 114 | 7 | 0.9 | 768  | 4 | US-09-252-991A-23008 | Sequence 23008, A  | 187 | 6 | 0.8 | 14 | 2 | US-08-929-414-15   | Sequence 15, Appl  |
| 115 | 7 | 0.9 | 822  | 4 | US-08-684-932A-38    | Sequence 38, Appl  | 188 | 6 | 0.8 | 14 | 3 | US-08-834-306-61   | Sequence 61, Appl  |
| 116 | 7 | 0.9 | 904  | 3 | US-08-632-537-2      | Sequence 2, Appli  | 189 | 6 | 0.8 | 14 | 3 | US-08-834-306-62   | Sequence 62, Appl  |
| 117 | 7 | 0.9 | 904  | 4 | US-09-894-998A-18    | Sequence 18, Appl  | 190 | 6 | 0.8 | 14 | 3 | US-08-993-674A-61  | Sequence 61, Appl  |
| 118 | 7 | 0.9 | 904  | 5 | PCT-US96-05316-2     | Sequence 2, Appli  | 191 | 6 | 0.8 | 14 | 3 | US-08-993-674A-62  | Sequence 62, Appl  |
| 119 | 7 | 0.9 | 904  | 6 | 5244792-3            | Patent No. 5244792 | 192 | 6 | 0.8 | 14 | 4 | US-09-256-976-61   | Sequence 61, Appl  |
| 120 | 7 | 0.9 | 930  | 4 | US-09-254-594-6      | Sequence 6, Appli  | 193 | 6 | 0.8 | 14 | 4 | US-09-256-976-62   | Sequence 62, Appl  |
| 121 | 7 | 0.9 | 984  | 4 | US-09-328-352-6926   | Sequence 6926, Ap  | 194 | 6 | 0.8 | 14 | 4 | US-08-766-596A-39  | Sequence 39, Appl  |
| 122 | 7 | 0.9 | 1005 | 4 | US-09-770-170-4      | Sequence 4, Appli  | 195 | 6 | 0.8 | 15 | 1 | US-08-320-373-17   | Sequence 17, Appl  |
| 123 | 7 | 0.9 | 1043 | 3 | US-08-928-361B-30    | Sequence 30, Appl  | 196 | 6 | 0.8 | 15 | 1 | US-08-320-373-18   | Sequence 18, Appl  |
| 124 | 7 | 0.9 | 1095 | 4 | US-09-206-942-69     | Sequence 69, Appl  | 197 | 6 | 0.8 | 15 | 1 | US-08-320-373-63   | Sequence 63, Appl  |
| 125 | 7 | 0.9 | 1151 | 4 | US-09-328-352-4744   | Sequence 4744, Ap  | 198 | 6 | 0.8 | 15 | 1 | US-08-320-373-64   | Sequence 64, Appl  |
| 126 | 7 | 0.9 | 1279 | 4 | US-09-724-517-2      | Sequence 2, Appli  | 199 | 6 | 0.8 | 15 | 1 | US-08-320-373-65   | Sequence 65, Appl  |
| 127 | 7 | 0.9 | 1279 | 4 | US-09-641-807A-2     | Sequence 2, Appli  | 200 | 6 | 0.8 | 15 | 1 | US-08-320-373-66   | Sequence 66, Appl  |
| 128 | 7 | 0.9 | 1279 | 4 | US-09-723-096-2      | Sequence 2, Appli  | 201 | 6 | 0.8 | 15 | 1 | US-08-320-373-67   | Sequence 67, Appl  |
| 129 | 7 | 0.9 | 1520 | 4 | US-09-252-991A-21336 | Sequence 21336, A  | 202 | 6 | 0.8 | 15 | 1 | US-08-320-373-68   | Sequence 68, Appl  |
| 130 | 7 | 0.9 | 1536 | 1 | US-08-038-682-2      | Sequence 2, Appli  | 203 | 6 | 0.8 | 15 | 2 | US-08-320-373-69   | Sequence 69, Appl  |
| 131 | 7 | 0.9 | 1536 | 2 | US-08-302-832-2      | Sequence 2, Appli  | 204 | 6 | 0.8 | 15 | 2 | US-08-265-310-17   | Sequence 17, Appl  |
| 132 | 7 | 0.9 | 1536 | 2 | US-08-530-198-2      | Sequence 2, Appli  | 205 | 6 | 0.8 | 15 | 3 | US-08-951-742-17   | Sequence 17, Appl  |
| 133 | 7 | 0.9 | 1536 | 2 | US-08-468-880-2      | Sequence 2, Appli  | 206 | 6 | 0.8 | 15 | 4 | US-08-766-596A-38  | Sequence 38, Appl  |
| 134 | 7 | 0.9 | 1536 | 2 | US-08-728-470-2      | Sequence 2, Appli  | 207 | 6 | 0.8 | 16 | 1 | US-08-009-448-3    | Sequence 3, Appli  |
| 135 | 7 | 0.9 | 1536 | 3 | US-08-617-697-2      | Sequence 2, Appli  | 208 | 6 | 0.8 | 16 | 1 | US-08-320-373-83   | Sequence 83, Appli |
| 136 | 7 | 0.9 | 1536 | 3 | US-08-719-641-2      | Sequence 2, Appli  | 209 | 6 | 0.8 | 16 | 1 | US-08-320-373-83   | Sequence 83, Appli |
| 137 | 7 | 0.9 | 1536 | 7 | US-09-206-942-67     | Sequence 67, Appl  | 210 | 6 | 0.8 | 16 | 1 | US-08-400-607-3    | Sequence 3, Appli  |
| 138 | 7 | 0.9 | 1721 | 3 | US-08-700-651-5      | Sequence 5, Appli  | 211 | 6 | 0.8 | 16 | 3 | US-08-462-436-19   | Sequence 19, Appl  |
| 139 | 7 | 0.9 | 1721 | 3 | US-08-928-361B-6     | Sequence 6, Appli  | 212 | 6 | 0.8 | 16 | 3 | US-08-462-436-20   | Sequence 20, Appl  |
| 140 | 7 | 0.9 | 1721 | 3 | US-09-588-995A-4     | Sequence 6, Appli  | 213 | 6 | 0.8 | 16 | 3 | US-08-463-275-19   | Sequence 19, Appl  |
| 141 | 7 | 0.9 | 3170 | 3 | US-09-036-987A-6     | Sequence 4, Appli  | 214 | 6 | 0.8 | 16 | 3 | US-08-463-275-20   | Sequence 20, Appl  |
| 142 | 7 | 0.9 | 3170 | 3 | US-09-370-700-4      | Sequence 4, Appli  | 215 | 6 | 0.8 | 16 | 3 | US-08-736-915-3    | Sequence 3, Appli  |
| 143 | 7 | 0.9 | 3170 | 4 | US-09-603-207-4      | Sequence 4, Appli  | 216 | 6 | 0.8 | 16 | 4 | US-08-640-877-19   | Sequence 19, Appl  |
| 144 | 7 | 0.9 | 3562 | 4 | US-09-679-279-14     | Sequence 14, Appl  | 217 | 6 | 0.8 | 16 | 4 | US-08-640-877-20   | Sequence 20, Appl  |
| 145 | 6 | 0.8 | 6    | 2 | US-08-265-310-14     | Sequence 14, Appl  | 218 | 6 | 0.8 | 16 | 4 | US-09-799-576A-19  | Sequence 19, Appl  |
| 146 | 6 | 0.8 | 6    | 3 | US-08-951-742-14     | Sequence 14, Appl  | 219 | 6 | 0.8 | 16 | 4 | US-09-799-576A-20  | Sequence 20, Appl  |
| 147 | 6 | 0.8 | 6    | 4 | US-09-187-859-2549   | Sequence 2549, Ap  | 220 | 6 | 0.8 | 16 | 4 | US-09-799-576A-20  | Sequence 20, Appl  |
| 148 | 6 | 0.8 | 8    | 4 | US-09-839-542B-2549  | Sequence 2549, Ap  | 221 | 6 | 0.8 | 18 | 5 | PCT-US93-06751-139 | Patent No. 5182210 |
| 149 | 6 | 0.8 | 8    | 4 | US-09-187-859-2430   | Sequence 2430, Ap  | 222 | 6 | 0.8 | 20 | 6 | 5182210-4          | Sequence 3, Appli  |
| 150 | 6 | 0.8 | 8    | 4 | US-09-187-859-2432   | Sequence 2432, Ap  | 223 | 6 | 0.8 | 21 | 1 | US-08-403-379A-3   | Sequence 3, Appli  |
| 151 | 6 | 0.8 | 8    | 4 | US-09-187-859-2550   | Sequence 2550, Ap  | 224 | 6 | 0.8 | 21 | 2 | US-08-557-309B-52  | Sequence 52, Appl  |
| 152 | 6 | 0.8 | 8    | 4 | US-09-839-542B-2430  | Sequence 2430, Ap  | 225 | 6 | 0.8 | 21 | 2 | US-08-557-309B-57  | Sequence 57, Appl  |
| 153 | 6 | 0.8 | 8    | 4 | US-09-839-542B-2432  | Sequence 2432, Ap  | 226 | 6 | 0.8 | 21 | 2 | US-08-557-309B-58  | Sequence 58, Appl  |
| 154 | 6 | 0.8 | 8    | 4 | US-09-839-542B-2550  | Sequence 2550, Ap  | 227 | 6 | 0.8 | 21 | 2 | US-08-929-414-3    | Sequence 3, Appli  |
| 155 | 6 | 0.8 | 9    | 1 | US-08-320-373-23     | Sequence 23, Appl  | 228 | 6 | 0.8 | 21 | 2 | US-08-929-414-10   | Sequence 10, Appl  |
| 156 | 6 | 0.8 | 9    | 1 | US-08-320-373-28     | Sequence 28, Appl  | 229 | 6 | 0.8 | 21 | 2 | US-08-929-414-11   | Sequence 11, Appl  |
| 157 | 6 | 0.8 | 9    | 4 | US-09-187-859-2431   | Sequence 2431, Ap  | 230 | 6 | 0.8 | 21 | 2 | US-08-997-080-13   | Sequence 13, Appl  |
| 158 | 6 | 0.8 | 9    | 4 | US-09-187-859-2435   | Sequence 2435, Ap  | 231 | 6 | 0.8 | 21 | 2 | US-08-997-080-13   | Sequence 13, Appl  |
| 159 | 6 | 0.8 | 9    | 4 | US-09-187-859-2551   | Sequence 2551, Ap  | 232 | 6 | 0.8 | 21 | 3 | US-08-873-970-13   | Sequence 13, Appl  |
| 160 | 6 | 0.8 | 9    | 4 | US-08-766-596A-43    | Sequence 43, Appl  | 233 | 6 | 0.8 | 21 | 3 | US-08-834-306-55   | Sequence 55, Appl  |
| 161 | 6 | 0.8 | 9    | 4 | US-09-839-542B-2431  | Sequence 2431, Ap  | 234 | 6 | 0.8 | 21 | 3 | US-08-834-306-56   | Sequence 56, Appl  |
| 162 | 6 | 0.8 | 9    | 4 | US-09-839-542B-2435  | Sequence 2435, Ap  | 235 | 6 | 0.8 | 21 | 3 | US-08-834-306-58   | Sequence 58, Appl  |
| 163 | 6 | 0.8 | 9    | 4 | US-09-839-542B-2551  | Sequence 2551, Ap  | 236 | 6 | 0.8 | 21 | 3 | US-09-095-855-13   | Sequence 13, Appl  |
| 164 | 6 | 0.8 | 10   | 1 | US-08-320-373-5      | Sequence 5, Appli  | 237 | 6 | 0.8 | 21 | 3 | US-08-993-674A-55  | Sequence 55, Appl  |
| 165 | 6 | 0.8 | 10   | 1 | US-08-320-373-22     | Sequence 22, Appl  | 238 | 6 | 0.8 | 21 | 3 | US-08-993-674A-56  | Sequence 56, Appl  |
| 166 | 6 | 0.8 | 10   | 1 | US-08-320-373-29     | Sequence 29, Appl  | 239 | 6 | 0.8 | 21 | 3 | US-08-993-674A-58  | Sequence 58, Appl  |
| 167 | 6 | 0.8 | 10   | 1 | US-08-707-860D-2     | Sequence 2, Appli  | 240 | 6 | 0.8 | 21 | 3 | US-09-248-588-46   | Sequence 46, Appl  |
| 168 | 6 | 0.8 | 10   | 4 | US-08-766-596A-42    | Sequence 42, Appl  | 241 | 6 | 0.8 | 21 | 4 | US-09-324-542-13   | Sequence 13, Appl  |
| 169 | 6 | 0.8 | 10   | 4 | US-08-320-373-25     | Sequence 25, Appl  | 242 | 6 | 0.8 | 21 | 4 | US-09-205-426-13   | Sequence 13, Appl  |
| 170 | 6 | 0.8 | 11   | 3 | US-09-208-966-41     | Sequence 41, Appl  | 243 | 6 | 0.8 | 21 | 4 | US-09-205-426-13   | Sequence 13, Appl  |
| 171 | 6 | 0.8 | 11   | 4 | US-08-707-860D-4     | Sequence 4, Appli  | 244 | 6 | 0.8 | 21 | 4 | US-09-256-976-55   | Sequence 55, Appl  |
| 172 | 6 | 0.8 | 11   | 4 | US-09-555-352-21     | Sequence 21, Appl  | 245 | 6 | 0.8 | 21 | 4 | US-09-256-976-56   | Sequence 56, Appl  |
| 173 | 6 | 0.8 | 11   | 4 | US-09-555-352-21     | Sequence 21, Appl  | 246 | 6 | 0.8 | 21 | 4 | US-09-256-976-58   | Sequence 58, Appl  |

247 6 0.8 23 3 US-08-609-236-3 Sequence 3, Appli  
248 6 0.8 23 4 US-09-431-705-15 Sequence 15, Appli  
249 6 0.8 25 3 US-08-894-699-81 Sequence 81, Appli  
250 6 0.8 25 3 US-09-444-410-81 Sequence 81, Appli  
251 6 0.8 26 6 5208158-11 Patent No. 5208158  
252 6 0.8 27 4 US-07-946-180B-4 Sequence 4, Appli  
253 6 0.8 31 3 US-08-822-324-12 Sequence 12, Appli  
254 6 0.8 32 1 US-08-009-448-1 Sequence 1, Appli  
255 6 0.8 32 1 US-08-400-607-1 Sequence 1, Appli  
256 6 0.8 33 4 US-09-149-476-660 Sequence 660, Appli  
257 6 0.8 36 4 US-08-817-441-85 Sequence 85, Appli  
258 6 0.8 37 1 US-08-211-202-120 Sequence 120, Appli  
259 6 0.8 38 1 US-08-211-202-121 Sequence 121, Appli  
260 6 0.8 39 6 5168051-5 Patent No. 5168051  
261 6 0.8 42 4 US-08-481-968A-12 Sequence 12, Appli  
262 6 0.8 42 4 US-08-481-968A-13 Sequence 13, Appli  
263 6 0.8 42 4 US-08-154-712B-12 Sequence 12, Appli  
264 6 0.8 42 4 US-08-154-712B-13 Sequence 13, Appli  
265 6 0.8 49 4 US-08-916-246-11 Sequence 11, Appli  
266 6 0.8 51 6 5168051-4 Patent No. 5168051  
267 6 0.8 57 4 US-08-122-570-3 Sequence 3, Appli  
268 6 0.8 57 4 US-08-932-411A-3 Sequence 3, Appli  
269 6 0.8 59 4 US-09-450-072-77 Sequence 77, Appli  
270 6 0.8 60 4 US-09-351-348-77 Sequence 77, Appli  
271 6 0.8 65 4 US-09-363-939A-122 Sequence 122, Appli  
272 6 0.8 65 4 US-09-134-001C-5364 Sequence 5364, Appli  
273 6 0.8 69 4 US-09-134-001C-4158 Sequence 4158, Appli  
274 6 0.8 69 4 US-09-916-510A-6 Sequence 6, Appli  
275 6 0.8 69 4 US-09-328-352-7903 Sequence 7903, Appli  
276 6 0.8 69 4 US-09-107-532A-5849 Sequence 5849, Appli  
277 6 0.8 72 1 US-08-280-443-8 Sequence 8, Appli  
278 6 0.8 72 1 US-08-457-459-8 Sequence 8, Appli  
279 6 0.8 72 1 US-08-555-678-8 Sequence 8, Appli  
280 6 0.8 72 5 PCT-US98-03275-8 Sequence 24, Appli  
281 6 0.8 73 4 US-09-366-887A-24 Sequence 24, Appli  
282 6 0.8 73 4 US-09-328-352-7260 Sequence 7260, Appli  
283 6 0.8 74 4 US-09-072-596-286 Sequence 286, Appli  
284 6 0.8 76 4 US-09-134-001C-3170 Sequence 3170, Appli  
285 6 0.8 78 4 US-09-732-210-1006 Sequence 1006, Appli  
286 6 0.8 82 4 US-09-107-532A-4748 Sequence 4748, Appli  
287 6 0.8 83 3 US-08-834-306-63 Sequence 63, Appli  
288 6 0.8 83 3 US-08-993-674A-63 Sequence 63, Appli  
289 6 0.8 84 4 US-09-256-976-63 Sequence 63, Appli  
290 6 0.8 84 4 US-09-107-532A-5149 Sequence 5149, Appli  
291 6 0.8 85 4 US-09-476-366A-2 Sequence 2, Appli  
292 6 0.8 85 4 US-09-256-976-62 Sequence 82, Appli  
293 6 0.8 87 4 US-09-107-532A-5431 Sequence 5431, Appli  
294 6 0.8 87 4 US-09-107-532A-5948 Sequence 5948, Appli  
295 6 0.8 89 4 US-09-194-488A-28 Sequence 28, Appli  
296 6 0.8 90 4 US-09-252-991A-25974 Sequence 25974, Appli  
297 6 0.8 93 3 US-08-894-699-63 Sequence 63, Appli  
298 6 0.8 93 3 US-09-444-410-63 Sequence 63, Appli  
299 6 0.8 94 1 US-08-450-945-63 Sequence 63, Appli  
300 6 0.8 94 3 US-08-976-161-63 Sequence 63, Appli  
301 6 0.8 94 4 US-09-256-976-95 Sequence 95, Appli  
302 6 0.8 96 4 US-09-732-210-1618 Sequence 1618, Appli  
303 6 0.8 98 3 US-08-478-097A-1 Sequence 1, Appli  
304 6 0.8 98 3 US-08-478-097A-1 Sequence 1, Appli  
305 6 0.8 98 3 US-08-931-856B-150 Sequence 150, Appli  
306 6 0.8 98 3 US-08-981-739-150 Sequence 150, Appli  
307 6 0.8 98 4 US-09-128-026-150 Sequence 1, Appli  
308 6 0.8 98 4 US-09-496-398-1 Sequence 4, Appli  
309 6 0.8 98 4 US-09-252-991A-32409 Sequence 32409, Appli  
310 6 0.8 101 4 US-09-552-142A-13 Sequence 13, Appli  
311 6 0.8 103 1 US-08-707-860D-9 Sequence 9, Appli  
312 6 0.8 103 4 US-08-707-860D-11 Sequence 11, Appli  
313 6 0.8 105 4 US-08-311-731A-387 Sequence 387, Appli  
314 6 0.8 107 3 US-09-274-642-38 Sequence 38, Appli  
315 6 0.8 109 3 US-09-274-642-32 Sequence 32, Appli  
316 6 0.8 109 3 US-09-274-642-34 Sequence 34, Appli  
317 6 0.8 109 4 US-09-252-991A-24976 Sequence 24976, Appli  
318 6 0.8 110 4 US-09-134-001C-3334 Sequence 3334, Appli

320 6 0.8 110 4 US-09-462-843A-3 Sequence 3, Appli  
321 6 0.8 110 4 US-09-904-615-149 Sequence 149, Appli  
322 6 0.8 111 3 US-09-274-642-39 Sequence 39, Appli  
323 6 0.8 111 4 US-08-916-246-2 Sequence 2, Appli  
324 6 0.8 112 1 US-07-979-441-1 Sequence 1, Appli  
325 6 0.8 112 1 US-08-197-792-36 Sequence 36, Appli  
326 6 0.8 112 1 US-08-486-057B-41 Sequence 41, Appli  
327 6 0.8 112 1 US-08-459-850-36 Sequence 36, Appli  
328 6 0.8 112 1 US-08-459-214-36 Sequence 36, Appli  
329 6 0.8 112 1 US-08-470-837-30 Sequence 30, Appli  
330 6 0.8 112 2 US-08-789-588-41 Sequence 41, Appli  
331 6 0.8 112 2 US-08-410-573-1 Sequence 1, Appli  
332 6 0.8 112 3 US-09-123-233-2 Sequence 2, Appli  
333 6 0.8 112 3 US-09-123-233-8 Sequence 8, Appli  
334 6 0.8 112 3 US-08-927-433-5 Sequence 5, Appli  
335 6 0.8 112 4 US-08-868-452-30 Sequence 30, Appli  
336 6 0.8 112 4 US-09-095-637D-1 Sequence 1, Appli  
337 6 0.8 112 5 PCT-US93-03068-1 Sequence 1, Appli  
338 6 0.8 113 1 US-08-241-853-15 Sequence 15, Appli  
339 6 0.8 113 2 US-08-850-917-15 Sequence 15, Appli  
340 6 0.8 114 1 US-08-481-377-23 Sequence 23, Appli  
341 6 0.8 114 2 US-08-491-835-21 Sequence 21, Appli  
342 6 0.8 114 3 US-09-153-733A-23 Sequence 23, Appli  
343 6 0.8 114 3 US-08-946-092A-21 Sequence 21, Appli  
344 6 0.8 114 3 US-08-927-433-6 Sequence 6, Appli  
345 6 0.8 114 3 US-09-172-082-21 Sequence 21, Appli  
346 6 0.8 114 3 US-08-624-635-22 Sequence 22, Appli  
347 6 0.8 114 4 US-09-301-520D-21 Sequence 21, Appli  
348 6 0.8 114 4 US-09-389-705-23 Sequence 23, Appli  
349 6 0.8 114 5 PCT-US94-00665-23 Sequence 23, Appli  
350 6 0.8 114 5 PCT-US94-00665-21 Sequence 21, Appli  
351 6 0.8 115 1 US-08-551-539B-19 Sequence 19, Appli  
352 6 0.8 115 1 US-08-455-559-25 Sequence 25, Appli  
353 6 0.8 115 2 US-08-525-596B-29 Sequence 29, Appli  
354 6 0.8 115 2 US-08-581-528A-19 Sequence 19, Appli  
355 6 0.8 115 3 US-09-097-616-19 Sequence 19, Appli  
356 6 0.8 115 3 US-09-145-060-25 Sequence 25, Appli  
357 6 0.8 115 3 US-09-378-238-39 Sequence 39, Appli  
358 6 0.8 115 4 US-09-629-938-29 Sequence 29, Appli  
359 6 0.8 115 5 PCT-US94-00657-25 Sequence 25, Appli  
360 6 0.8 115 5 PCT-US94-07762-19 Sequence 19, Appli  
361 6 0.8 115 5 PCT-US94-07789-19 Sequence 19, Appli  
362 6 0.8 116 1 US-08-481-377-26 Sequence 26, Appli  
363 6 0.8 116 2 US-08-491-835-24 Sequence 24, Appli  
364 6 0.8 116 3 US-09-153-733A-26 Sequence 26, Appli  
365 6 0.8 116 3 US-08-946-092A-24 Sequence 24, Appli  
366 6 0.8 116 3 US-09-172-062-24 Sequence 24, Appli  
367 6 0.8 116 4 US-09-301-520D-24 Sequence 24, Appli  
368 6 0.8 116 4 US-09-389-705-26 Sequence 26, Appli  
369 6 0.8 116 5 PCT-US94-00665-26 Sequence 26, Appli  
370 6 0.8 116 5 PCT-US94-00665-24 Sequence 24, Appli  
371 6 0.8 117 4 US-09-702-705-1673 Sequence 1673, Appli  
372 6 0.8 117 4 US-09-736-457-1673 Sequence 1673, Appli  
373 6 0.8 117 4 US-09-134-001C-4174 Sequence 4174, Appli  
374 6 0.8 118 4 US-09-252-991A-18728 Sequence 18728, Appli  
375 6 0.8 118 4 US-09-732-210-1109 Sequence 1109, Appli  
376 6 0.8 119 4 US-09-252-991A-22272 Sequence 22272, Appli  
377 6 0.8 120 4 US-09-252-991A-25468 Sequence 25468, Appli  
378 6 0.8 121 4 US-09-194-468A-27 Sequence 27, Appli  
379 6 0.8 122 4 US-09-328-352-4619 Sequence 4619, Appli  
380 6 0.8 122 4 US-09-134-001C-4957 Sequence 4957, Appli  
381 6 0.8 124 4 US-09-107-532A-4998 Sequence 4998, Appli  
382 6 0.8 126 4 US-09-328-352-5607 Sequence 5607, Appli  
383 6 0.8 127 4 US-09-107-532A-6990 Sequence 6990, Appli  
384 6 0.8 129 4 US-09-252-991A-24453 Sequence 24453, Appli  
385 6 0.8 130 4 US-08-728-742A-8 Sequence 8, Appli  
386 6 0.8 130 4 US-08-203-005B-10 Sequence 10, Appli  
387 6 0.8 135 1 US-09-352-991A-2652 Sequence 2652, Appli  
388 6 0.8 137 4 US-09-252-991A-19622 Sequence 19622, Appli  
389 6 0.8 138 4 US-08-065-844A-8 Sequence 8, Appli  
390 6 0.8 139 4 US-08-691-071-4 Sequence 4, Appli  
391 6 0.8 140 1 US-08-843-370-4 Sequence 4, Appli  
392 6 0.8 140 2 US-08-843-370-4 Sequence 4, Appli

|     |   |     |     |   |                      |
|-----|---|-----|-----|---|----------------------|
| 333 | 6 | 0.8 | 141 | 4 | US-09-134-001C-4183  |
| 334 | 6 | 0.8 | 141 | 4 | US-09-252-991A-23722 |
| 335 | 6 | 0.8 | 142 | 4 | US-09-252-991A-19398 |
| 336 | 6 | 0.8 | 143 | 4 | US-09-198-452A-6     |
| 337 | 6 | 0.8 | 144 | 4 | US-09-252-991A-19305 |
| 338 | 6 | 0.8 | 145 | 2 | US-08-997-080-55     |
| 339 | 6 | 0.8 | 145 | 2 | US-08-997-362-55     |
| 340 | 6 | 0.8 | 145 | 3 | US-08-873-970-55     |
| 341 | 6 | 0.8 | 145 | 3 | US-08-959-855-55     |
| 342 | 6 | 0.8 | 145 | 3 | US-08-705-347A-55    |
| 343 | 6 | 0.8 | 145 | 4 | US-09-324-542-55     |
| 344 | 6 | 0.8 | 145 | 4 | US-09-205-426-55     |
| 345 | 6 | 0.8 | 145 | 4 | US-09-200-643-55     |
| 346 | 6 | 0.8 | 145 | 4 | US-09-702-705-1672   |
| 347 | 6 | 0.8 | 145 | 4 | US-09-736-457-1672   |
| 348 | 6 | 0.8 | 145 | 4 | US-08-252-991A-22680 |
| 349 | 6 | 0.8 | 145 | 4 | US-09-252-991A-23754 |
| 350 | 6 | 0.8 | 145 | 4 | US-09-252-991A-23264 |
| 351 | 6 | 0.8 | 146 | 4 | US-09-306-446C-5     |
| 352 | 6 | 0.8 | 147 | 4 | US-08-858-207A-419   |
| 353 | 6 | 0.8 | 147 | 4 | US-09-252-991A-21279 |
| 354 | 6 | 0.8 | 147 | 4 | US-09-252-991A-17720 |
| 355 | 6 | 0.8 | 150 | 4 | US-08-252-991A-25932 |
| 356 | 6 | 0.8 | 150 | 4 | US-09-732-210-620    |
| 357 | 6 | 0.8 | 151 | 4 | US-09-252-991A-31495 |
| 358 | 6 | 0.8 | 151 | 4 | US-09-732-210-622    |
| 359 | 6 | 0.8 | 151 | 4 | US-09-252-991A-18469 |
| 360 | 6 | 0.8 | 152 | 4 | US-09-352-991A-31944 |
| 361 | 6 | 0.8 | 155 | 4 | US-09-134-001C-4770  |
| 362 | 6 | 0.8 | 155 | 4 | US-09-252-991A-17807 |
| 363 | 6 | 0.8 | 155 | 4 | US-09-252-991A-25452 |
| 364 | 6 | 0.8 | 156 | 4 | US-09-252-991A-24611 |
| 365 | 6 | 0.8 | 157 | 4 | US-09-732-210-769    |
| 366 | 6 | 0.8 | 157 | 4 | US-09-732-210-1301   |
| 367 | 6 | 0.8 | 158 | 4 | US-09-134-001C-4336  |
| 368 | 6 | 0.8 | 159 | 4 | US-09-252-991A-18088 |
| 369 | 6 | 0.8 | 161 | 2 | US-08-403-852D-22    |
| 370 | 6 | 0.8 | 161 | 3 | US-09-510-646B-23    |
| 371 | 6 | 0.8 | 161 | 3 | US-09-231-918-22     |
| 372 | 6 | 0.8 | 161 | 4 | US-09-461-325-210    |
| 373 | 6 | 0.8 | 162 | 4 | US-09-252-991A-22151 |
| 374 | 6 | 0.8 | 162 | 4 | US-09-328-352-5454   |
| 375 | 6 | 0.8 | 163 | 4 | US-09-328-352-5791   |
| 376 | 6 | 0.8 | 164 | 2 | US-08-918-723-3      |
| 377 | 6 | 0.8 | 164 | 2 | US-09-237-507-3      |
| 378 | 6 | 0.8 | 165 | 1 | US-08-051-142-2      |
| 379 | 6 | 0.8 | 165 | 4 | US-09-232-991A-31171 |
| 380 | 6 | 0.8 | 166 | 4 | US-09-252-991A-19652 |
| 381 | 6 | 0.8 | 169 | 3 | US-08-833-167-58     |
| 382 | 6 | 0.8 | 169 | 4 | US-09-344-837A-58    |
| 383 | 6 | 0.8 | 169 | 4 | US-09-252-991A-22543 |
| 384 | 6 | 0.8 | 169 | 4 | US-09-252-991A-26040 |
| 385 | 6 | 0.8 | 169 | 4 | US-09-252-991A-29056 |
| 386 | 6 | 0.8 | 170 | 4 | US-09-198-452A-95    |
| 387 | 6 | 0.8 | 171 | 3 | US-08-833-167-33     |
| 388 | 6 | 0.8 | 171 | 3 | US-08-833-167-44     |
| 389 | 6 | 0.8 | 171 | 3 | US-08-833-167-45     |
| 390 | 6 | 0.8 | 171 | 3 | US-08-833-167-47     |
| 391 | 6 | 0.8 | 171 | 3 | US-08-833-167-57     |
| 392 | 6 | 0.8 | 171 | 3 | US-08-833-167-59     |
| 393 | 6 | 0.8 | 171 | 3 | US-08-833-167-60     |
| 394 | 6 | 0.8 | 171 | 4 | US-09-344-837A-43    |
| 395 | 6 | 0.8 | 171 | 4 | US-09-344-837A-44    |
| 396 | 6 | 0.8 | 171 | 4 | US-09-344-837A-45    |
| 397 | 6 | 0.8 | 171 | 4 | US-09-344-837A-47    |
| 398 | 6 | 0.8 | 171 | 4 | US-09-344-837A-57    |
| 399 | 6 | 0.8 | 171 | 4 | US-09-344-837A-59    |
| 400 | 6 | 0.8 | 171 | 4 | US-09-344-837A-60    |
| 401 | 6 | 0.8 | 173 | 1 | US-08-354-456A-7     |
| 402 | 6 | 0.8 | 173 | 1 | US-08-354-456A-7     |
| 403 | 6 | 0.8 | 173 | 4 | US-09-252-991A-23777 |
| 404 | 6 | 0.8 | 173 | 4 | US-09-252-991A-25321 |
| 405 | 6 | 0.8 | 174 | 1 | US-07-734-245A-43    |
| 406 | 6 | 0.8 | 174 | 1 | US-07-692-995B-46    |
| 407 | 6 | 0.8 | 174 | 1 | US-07-692-995B-46    |

|                   |     |   |     |     |   |                    |                   |
|-------------------|-----|---|-----|-----|---|--------------------|-------------------|
| Sequence 4183, Ap | 466 | 6 | 0.8 | 174 | 1 | US-08-010-099-82   | Sequence 82, Appl |
| Sequence 23722, A | 467 | 6 | 0.8 | 174 | 1 | US-08-225-224-5    | Sequence 5, Appl  |
| Sequence 19398, A | 468 | 6 | 0.8 | 174 | 1 | US-08-434-411-2    | Sequence 2, Appl  |
| Sequence 6, Appl  | 469 | 6 | 0.8 | 174 | 1 | US-08-434-402-2    | Sequence 2, Appl  |
| Sequence 19305, A | 470 | 6 | 0.8 | 174 | 1 | US-08-488-457-43   | Sequence 43, Appl |
| Sequence 55, Appl | 471 | 6 | 0.8 | 174 | 1 | US-08-488-716-82   | Sequence 82, Appl |
| Sequence 55, Appl | 472 | 6 | 0.8 | 174 | 1 | US-08-783-288-2    | Sequence 2, Appl  |
| Sequence 55, Appl | 473 | 6 | 0.8 | 174 | 1 | US-08-431-459A-31  | Sequence 31, Appl |
| Sequence 55, Appl | 474 | 6 | 0.8 | 174 | 2 | US-08-690-640-2    | Sequence 2, Appl  |
| Sequence 55, Appl | 475 | 6 | 0.8 | 174 | 3 | US-08-732-258-5    | Sequence 5, Appl  |
| Sequence 55, Appl | 476 | 6 | 0.8 | 174 | 3 | US-08-833-167-48   | Sequence 48, Appl |
| Sequence 55, Appl | 477 | 6 | 0.8 | 174 | 3 | US-08-833-167-49   | Sequence 49, Appl |
| Sequence 55, Appl | 478 | 6 | 0.8 | 174 | 3 | US-08-833-167-50   | Sequence 50, Appl |
| Sequence 1672, Ap | 479 | 6 | 0.8 | 174 | 3 | US-08-833-167-51   | Sequence 51, Appl |
| Sequence 1672, Ap | 480 | 6 | 0.8 | 174 | 3 | US-08-833-167-52   | Sequence 52, Appl |
| Sequence 22680, A | 481 | 6 | 0.8 | 174 | 3 | US-08-833-167-53   | Sequence 53, Appl |
| Sequence 27574, A | 482 | 6 | 0.8 | 174 | 3 | US-08-833-167-54   | Sequence 54, Appl |
| Sequence 29264, A | 483 | 6 | 0.8 | 174 | 3 | US-08-833-167-55   | Sequence 55, Appl |
| Sequence 5, Appl  | 484 | 6 | 0.8 | 174 | 3 | US-08-833-167-56   | Sequence 56, Appl |
| Sequence 419, App | 485 | 6 | 0.8 | 174 | 3 | US-08-833-167-95   | Sequence 95, Appl |
| Sequence 21279, A | 486 | 6 | 0.8 | 174 | 3 | US-08-833-167-97   | Sequence 97, Appl |
| Sequence 17720, A | 487 | 6 | 0.8 | 174 | 3 | US-08-833-167-98   | Sequence 98, Appl |
| Sequence 25932, A | 488 | 6 | 0.8 | 174 | 3 | US-08-833-167-99   | Sequence 99, Appl |
| Sequence 620, App | 489 | 6 | 0.8 | 174 | 3 | US-08-833-167-101  | Sequence 101, App |
| Sequence 31495, A | 490 | 6 | 0.8 | 174 | 3 | US-08-833-167-102  | Sequence 102, App |
| Sequence 622, App | 491 | 6 | 0.8 | 174 | 3 | US-08-833-167-103  | Sequence 103, App |
| Sequence 18469, A | 492 | 6 | 0.8 | 174 | 3 | US-09-221-181-1    | Sequence 1, Appl  |
| Sequence 31944, A | 493 | 6 | 0.8 | 174 | 3 | US-09-304-186-82   | Sequence 82, Appl |
| Sequence 4770, Ap | 494 | 6 | 0.8 | 174 | 4 | US-09-344-837A-48  | Sequence 48, Appl |
| Sequence 17807, A | 495 | 6 | 0.8 | 174 | 4 | US-09-344-837A-49  | Sequence 49, Appl |
| Sequence 25452, A | 496 | 6 | 0.8 | 174 | 4 | US-09-344-837A-50  | Sequence 50, Appl |
| Sequence 24611, A | 497 | 6 | 0.8 | 174 | 4 | US-09-344-837A-51  | Sequence 51, Appl |
| Sequence 769, App | 498 | 6 | 0.8 | 174 | 4 | US-09-344-837A-52  | Sequence 52, Appl |
| Sequence 1301, Ap | 499 | 6 | 0.8 | 174 | 4 | US-09-344-837A-53  | Sequence 53, Appl |
| Sequence 4336, Ap | 500 | 6 | 0.8 | 174 | 4 | US-09-344-837A-54  | Sequence 54, Appl |
| Sequence 22, Appl | 501 | 6 | 0.8 | 174 | 4 | US-09-344-837A-55  | Sequence 55, Appl |
| Sequence 18088, A | 502 | 6 | 0.8 | 174 | 4 | US-09-344-837A-56  | Sequence 56, Appl |
| Sequence 22, Appl | 503 | 6 | 0.8 | 174 | 4 | US-09-344-837A-95  | Sequence 95, Appl |
| Sequence 210, App | 504 | 6 | 0.8 | 174 | 4 | US-09-344-837A-96  | Sequence 96, Appl |
| Sequence 5454, Ap | 505 | 6 | 0.8 | 174 | 4 | US-09-344-837A-97  | Sequence 97, Appl |
| Sequence 5791, Ap | 506 | 6 | 0.8 | 174 | 4 | US-09-344-837A-98  | Sequence 98, Appl |
| Sequence 3, Appl  | 507 | 6 | 0.8 | 174 | 4 | US-09-344-837A-99  | Sequence 99, Appl |
| Sequence 3, Appl  | 508 | 6 | 0.8 | 174 | 4 | US-09-344-837A-101 | Sequence 101, App |
| Sequence 2, Appl  | 509 | 6 | 0.8 | 174 | 4 | US-09-344-837A-102 | Sequence 102, App |
| Sequence 31171, A | 510 | 6 | 0.8 | 174 | 4 | US-09-344-837A-103 | Sequence 103, App |
| Sequence 2, Appl  | 511 | 6 | 0.8 | 174 | 4 | US-09-904-196B-1   | Sequence 1, Appl  |
| Sequence 16562, A | 512 | 6 | 0.8 | 174 | 4 | US-09-198-452A-476 | Sequence 476, App |
| Sequence 58, Appl | 513 | 6 | 0.8 | 174 | 5 | PCT-US95-04468-5   | Sequence 5, Appl  |
| Sequence 58, Appl | 514 | 6 | 0.8 | 174 | 6 | 5194592-26         | Patent No. 5194   |

|     |   |     |     |   |                   |                    |     |   |     |     |   |                     |                     |
|-----|---|-----|-----|---|-------------------|--------------------|-----|---|-----|-----|---|---------------------|---------------------|
| 539 | 6 | 0.8 | 175 | 1 | US-08-010-099-89  | Sequence 89, Appl  | 512 | 6 | 0.8 | 175 | 3 | US-09-304-186-74    | Sequence 74, Appl   |
| 540 | 6 | 0.8 | 175 | 1 | US-08-010-099-90  | Sequence 90, Appl  | 513 | 6 | 0.8 | 175 | 3 | US-09-304-186-75    | Sequence 75, Appl   |
| 541 | 6 | 0.8 | 175 | 1 | US-08-010-099-91  | Sequence 91, Appl  | 514 | 6 | 0.8 | 175 | 3 | US-09-304-186-76    | Sequence 76, Appl   |
| 542 | 6 | 0.8 | 175 | 1 | US-08-010-099-92  | Sequence 92, Appl  | 515 | 6 | 0.8 | 175 | 3 | US-09-304-186-77    | Sequence 77, Appl   |
| 543 | 6 | 0.8 | 175 | 1 | US-08-010-099-93  | Sequence 93, Appl  | 516 | 6 | 0.8 | 175 | 3 | US-09-304-186-78    | Sequence 78, Appl   |
| 544 | 6 | 0.8 | 175 | 1 | US-08-010-099-94  | Sequence 94, Appl  | 517 | 6 | 0.8 | 175 | 3 | US-09-304-186-79    | Sequence 79, Appl   |
| 545 | 6 | 0.8 | 175 | 1 | US-08-010-099-95  | Sequence 95, Appl  | 518 | 6 | 0.8 | 175 | 3 | US-09-304-186-80    | Sequence 80, Appl   |
| 546 | 6 | 0.8 | 175 | 1 | US-08-010-099-96  | Sequence 96, Appl  | 519 | 6 | 0.8 | 175 | 3 | US-09-304-186-81    | Sequence 81, Appl   |
| 547 | 6 | 0.8 | 175 | 1 | US-08-010-099-97  | Sequence 97, Appl  | 520 | 6 | 0.8 | 175 | 3 | US-09-304-186-82    | Sequence 82, Appl   |
| 548 | 6 | 0.8 | 175 | 1 | US-08-010-099-98  | Sequence 98, Appl  | 521 | 6 | 0.8 | 175 | 3 | US-09-304-186-83    | Sequence 83, Appl   |
| 549 | 6 | 0.8 | 175 | 1 | US-08-010-099-99  | Sequence 99, Appl  | 522 | 6 | 0.8 | 175 | 3 | US-09-304-186-84    | Sequence 84, Appl   |
| 550 | 6 | 0.8 | 175 | 1 | US-08-010-099-100 | Sequence 100, Appl | 523 | 6 | 0.8 | 175 | 3 | US-09-304-186-85    | Sequence 85, Appl   |
| 551 | 6 | 0.8 | 175 | 1 | US-08-010-099-101 | Sequence 101, Appl | 524 | 6 | 0.8 | 175 | 3 | US-09-304-186-86    | Sequence 86, Appl   |
| 552 | 6 | 0.8 | 175 | 1 | US-08-010-099-102 | Sequence 102, Appl | 525 | 6 | 0.8 | 175 | 3 | US-09-304-186-87    | Sequence 87, Appl   |
| 553 | 6 | 0.8 | 175 | 1 | US-08-010-099-103 | Sequence 103, Appl | 526 | 6 | 0.8 | 175 | 3 | US-09-304-186-88    | Sequence 88, Appl   |
| 554 | 6 | 0.8 | 175 | 1 | US-08-010-099-104 | Sequence 104, Appl | 527 | 6 | 0.8 | 175 | 3 | US-09-304-186-89    | Sequence 89, Appl   |
| 555 | 6 | 0.8 | 175 | 1 | US-08-010-099-107 | Sequence 107, Appl | 528 | 6 | 0.8 | 175 | 3 | US-09-304-186-90    | Sequence 90, Appl   |
| 556 | 6 | 0.8 | 175 | 1 | US-08-010-099-109 | Sequence 109, Appl | 529 | 6 | 0.8 | 175 | 3 | US-09-304-186-91    | Sequence 91, Appl   |
| 557 | 6 | 0.8 | 175 | 1 | US-08-167-721-1   | Sequence 1, Appl   | 530 | 6 | 0.8 | 175 | 3 | US-09-304-186-92    | Sequence 92, Appl   |
| 558 | 6 | 0.8 | 175 | 1 | US-08-428-732-6   | Sequence 6, Appl   | 531 | 6 | 0.8 | 175 | 3 | US-09-304-186-93    | Sequence 93, Appl   |
| 559 | 6 | 0.8 | 175 | 1 | US-08-448-716-2   | Sequence 2, Appl   | 532 | 6 | 0.8 | 175 | 3 | US-09-304-186-94    | Sequence 94, Appl   |
| 560 | 6 | 0.8 | 175 | 1 | US-08-448-716-67  | Sequence 67, Appl  | 533 | 6 | 0.8 | 175 | 3 | US-09-304-186-95    | Sequence 95, Appl   |
| 561 | 6 | 0.8 | 175 | 1 | US-08-448-716-68  | Sequence 68, Appl  | 534 | 6 | 0.8 | 175 | 3 | US-09-304-186-96    | Sequence 96, Appl   |
| 562 | 6 | 0.8 | 175 | 1 | US-08-448-716-69  | Sequence 69, Appl  | 535 | 6 | 0.8 | 175 | 3 | US-09-304-186-97    | Sequence 97, Appl   |
| 563 | 6 | 0.8 | 175 | 1 | US-08-448-716-70  | Sequence 70, Appl  | 536 | 6 | 0.8 | 175 | 3 | US-09-304-186-98    | Sequence 98, Appl   |
| 564 | 6 | 0.8 | 175 | 1 | US-08-448-716-71  | Sequence 71, Appl  | 537 | 6 | 0.8 | 175 | 3 | US-09-304-186-99    | Sequence 99, Appl   |
| 565 | 6 | 0.8 | 175 | 1 | US-08-448-716-72  | Sequence 72, Appl  | 538 | 6 | 0.8 | 175 | 3 | US-09-304-186-100   | Sequence 100, Appl  |
| 566 | 6 | 0.8 | 175 | 1 | US-08-448-716-74  | Sequence 74, Appl  | 539 | 6 | 0.8 | 175 | 3 | US-09-304-186-101   | Sequence 101, Appl  |
| 567 | 6 | 0.8 | 175 | 1 | US-08-448-716-75  | Sequence 75, Appl  | 540 | 6 | 0.8 | 175 | 3 | US-09-304-186-102   | Sequence 102, Appl  |
| 568 | 6 | 0.8 | 175 | 1 | US-08-448-716-76  | Sequence 76, Appl  | 541 | 6 | 0.8 | 175 | 3 | US-09-304-186-103   | Sequence 103, Appl  |
| 569 | 6 | 0.8 | 175 | 1 | US-08-448-716-77  | Sequence 77, Appl  | 542 | 6 | 0.8 | 175 | 3 | US-09-304-186-104   | Sequence 104, Appl  |
| 570 | 6 | 0.8 | 175 | 1 | US-08-448-716-78  | Sequence 78, Appl  | 543 | 6 | 0.8 | 175 | 3 | US-09-304-186-107   | Sequence 107, Appl  |
| 571 | 6 | 0.8 | 175 | 1 | US-08-448-716-79  | Sequence 79, Appl  | 544 | 6 | 0.8 | 175 | 3 | US-09-304-186-109   | Sequence 109, Appl  |
| 572 | 6 | 0.8 | 175 | 1 | US-08-448-716-80  | Sequence 80, Appl  | 545 | 6 | 0.8 | 175 | 3 | US-09-230-733-1     | Sequence 1, Appl    |
| 573 | 6 | 0.8 | 175 | 1 | US-08-448-716-81  | Sequence 81, Appl  | 546 | 6 | 0.8 | 175 | 3 | PCT-US95-01729-2    | Sequence 2, Appl    |
| 574 | 6 | 0.8 | 175 | 1 | US-08-448-716-82  | Sequence 82, Appl  | 547 | 6 | 0.8 | 175 | 5 | US-08-031-148-2     | Sequence 2, Appl    |
| 575 | 6 | 0.8 | 175 | 1 | US-08-448-716-84  | Sequence 84, Appl  | 548 | 6 | 0.8 | 176 | 1 | US-08-031-148-2     | Sequence 2, Appl    |
| 576 | 6 | 0.8 | 175 | 1 | US-08-448-716-85  | Sequence 85, Appl  | 549 | 6 | 0.8 | 176 | 2 | US-08-306-078-1     | Sequence 1, Appl    |
| 577 | 6 | 0.8 | 175 | 1 | US-08-448-716-86  | Sequence 86, Appl  | 550 | 6 | 0.8 | 176 | 3 | US-08-415-838-2     | Sequence 2, Appl    |
| 578 | 6 | 0.8 | 175 | 1 | US-08-448-716-87  | Sequence 87, Appl  | 551 | 6 | 0.8 | 176 | 3 | US-08-463-318-2     | Sequence 161, Appl  |
| 579 | 6 | 0.8 | 175 | 1 | US-08-448-716-88  | Sequence 88, Appl  | 552 | 6 | 0.8 | 176 | 3 | US-08-463-318-161   | Sequence 162, Appl  |
| 580 | 6 | 0.8 | 175 | 1 | US-08-448-716-89  | Sequence 89, Appl  | 553 | 6 | 0.8 | 176 | 3 | US-08-468-609A-161  | Sequence 161, Appl  |
| 581 | 6 | 0.8 | 175 | 1 | US-08-448-716-89  | Sequence 89, Appl  | 554 | 6 | 0.8 | 176 | 3 | US-08-468-609A-162  | Sequence 162, Appl  |
| 582 | 6 | 0.8 | 175 | 1 | US-08-448-716-90  | Sequence 90, Appl  | 555 | 6 | 0.8 | 176 | 3 | US-08-149-101A-26   | Sequence 26, Appl   |
| 583 | 6 | 0.8 | 175 | 1 | US-08-448-716-90  | Sequence 90, Appl  | 556 | 6 | 0.8 | 176 | 3 | US-08-446-872A-161  | Sequence 161, Appl  |
| 584 | 6 | 0.8 | 175 | 1 | US-08-448-716-92  | Sequence 92, Appl  | 557 | 6 | 0.8 | 176 | 4 | US-08-446-872A-162  | Sequence 162, Appl  |
| 585 | 6 | 0.8 | 175 | 1 | US-08-448-716-93  | Sequence 93, Appl  | 558 | 6 | 0.8 | 176 | 4 | US-09-205-169-2     | Sequence 2, Appl    |
| 586 | 6 | 0.8 | 175 | 1 | US-08-448-716-94  | Sequence 94, Appl  | 559 | 6 | 0.8 | 176 | 4 | US-09-134-001C-3018 | Sequence 3018, Appl |
| 587 | 6 | 0.8 | 175 | 1 | US-08-448-716-95  | Sequence 95, Appl  | 560 | 6 | 0.8 | 176 | 4 | US-08-762-227A-161  | Sequence 161, Appl  |
| 588 | 6 | 0.8 | 175 | 1 | US-08-448-716-96  | Sequence 96, Appl  | 561 | 6 | 0.8 | 176 | 4 | US-08-762-227A-162  | Sequence 162, Appl  |
| 589 | 6 | 0.8 | 175 | 1 | US-08-448-716-97  | Sequence 97, Appl  | 562 | 6 | 0.8 | 176 | 5 | PCT-US94-12873-36   | Sequence 26, Appl   |
| 590 | 6 | 0.8 | 175 | 1 | US-08-448-716-98  | Sequence 98, Appl  | 563 | 6 | 0.8 | 176 | 5 | PCT-US95-01185-161  | Sequence 161, Appl  |
| 591 | 6 | 0.8 | 175 | 1 | US-08-448-716-99  | Sequence 99, Appl  | 564 | 6 | 0.8 | 176 | 5 | PCT-US95-01185-162  | Sequence 162, Appl  |
| 592 | 6 | 0.8 | 175 | 1 | US-08-448-716-100 | Sequence 100, Appl | 565 | 6 | 0.8 | 177 | 2 | US-08-338-793D-41   | Sequence 41, Appl   |
| 593 | 6 | 0.8 | 175 | 1 | US-08-448-716-101 | Sequence 101, Appl | 566 | 6 | 0.8 | 177 | 2 | US-08-338-793D-55   | Sequence 55, Appl   |
| 594 | 6 | 0.8 | 175 | 1 | US-08-448-716-102 | Sequence 102, Appl | 567 | 6 | 0.8 | 177 | 2 | US-08-431-459A-30   | Sequence 30, Appl   |
| 595 | 6 | 0.8 | 175 | 1 | US-08-448-716-103 | Sequence 103, Appl | 568 | 6 | 0.8 | 177 | 2 | US-08-797-689-14    | Sequence 14, Appl   |
| 596 | 6 | 0.8 | 175 | 1 | US-08-448-716-104 | Sequence 104, Appl | 569 | 6 | 0.8 | 177 | 3 | US-09-122-443-8     | Sequence 8, Appl    |
| 597 | 6 | 0.8 | 175 | 1 | US-08-448-716-107 | Sequence 107, Appl | 570 | 6 | 0.8 | 177 | 3 | US-08-833-167-116   | Sequence 116, Appl  |
| 598 | 6 | 0.8 | 175 | 1 | US-08-321-510-2   | Sequence 2, Appl   | 571 | 6 | 0.8 | 177 | 3 | US-08-833-167-117   | Sequence 117, Appl  |
| 599 | 6 | 0.8 | 175 | 2 | US-08-570-943-1   | Sequence 1, Appl   | 572 | 6 | 0.8 | 177 | 3 | US-08-833-167-118   | Sequence 118, Appl  |
| 600 | 6 | 0.8 | 175 | 2 | US-08-879-760-2   | Sequence 2, Appl   | 573 | 6 | 0.8 | 177 | 3 | US-08-833-167-119   | Sequence 119, Appl  |
| 601 | 6 | 0.8 | 175 | 2 | US-08-505-187-4   | Sequence 4, Appl   | 574 | 6 | 0.8 | 177 | 3 | US-08-833-167-120   | Sequence 120, Appl  |
| 602 | 6 | 0.8 | 175 | 3 | US-08-696-98-1    | Sequence 1, Appl   | 575 | 6 | 0.8 | 177 | 3 | US-08-833-167-121   | Sequence 121, Appl  |
| 603 | 6 | 0.8 | 175 | 3 | US-08-304-186-2   | Sequence 2, Appl   | 576 | 6 | 0.8 | 177 | 3 | US-08-833-167-122   | Sequence 122, Appl  |
| 604 | 6 | 0.8 | 175 | 3 | US-08-304-186-67  | Sequence 67, Appl  | 577 | 6 | 0.8 | 177 | 3 | US-08-833-167-123   | Sequence 123, Appl  |
| 605 | 6 | 0.8 | 175 | 3 | US-08-304-186-68  | Sequence 68, Appl  | 578 | 6 | 0.8 | 177 | 3 | US-08-833-167-124   | Sequence 124, Appl  |
| 606 | 6 | 0.8 | 175 | 3 | US-08-304-186-69  | Sequence 69, Appl  | 579 | 6 | 0.8 | 177 | 3 | US-08-833-167-125   | Sequence 125, Appl  |
| 607 | 6 | 0.8 | 175 | 3 | US-08-304-186-70  | Sequence 70, Appl  | 580 | 6 | 0.8 | 177 | 3 | US-08-833-167-126   | Sequence 126, Appl  |
| 608 | 6 | 0.8 | 175 | 3 | US-08-304-186-71  | Sequence 71, Appl  | 581 | 6 | 0.8 | 177 | 3 | US-08-833-167-127   | Sequence 127, Appl  |
| 609 | 6 | 0.8 | 175 | 3 | US-08-304-186-72  | Sequence 72, Appl  | 582 | 6 | 0.8 | 177 | 4 | US-09-344-837A-116  | Sequence 116, Appl  |
| 610 | 6 | 0.8 | 175 | 3 | US-09-304-186-73  | Sequence 73, Appl  | 583 | 6 | 0.8 | 177 | 4 | US-09-344-837A-117  | Sequence 117, Appl  |
| 611 | 6 | 0.8 | 175 | 3 | US-09-304-186-73  | Sequence 73, Appl  | 584 | 6 | 0.8 | 177 | 4 | US-09-344-837A-118  | Sequence 118, Appl  |

|     |   |     |     |   |                      |                    |     |     |   |                      |                    |
|-----|---|-----|-----|---|----------------------|--------------------|-----|-----|---|----------------------|--------------------|
| 695 | 6 | 0.8 | 177 | 4 | US-09-344-837A-119   | Sequence 119, App  | 758 | 217 | 4 | US-09-328-352-5784   | Sequence 5784, Ap  |
| 696 | 6 | 0.8 | 177 | 4 | US-09-344-837A-120   | Sequence 120, App  | 759 | 218 | 3 | US-09-107-532A-6250  | Sequence 6250, Ap  |
| 697 | 6 | 0.8 | 177 | 4 | US-09-344-837A-121   | Sequence 121, App  | 760 | 218 | 3 | US-09-216-001-4      | Sequence 4, Appl   |
| 698 | 6 | 0.8 | 177 | 4 | US-09-344-837A-122   | Sequence 122, App  | 761 | 218 | 3 | US-08-878-862-4      | Sequence 4, Appl   |
| 699 | 6 | 0.8 | 177 | 4 | US-09-344-837A-123   | Sequence 123, App  | 762 | 219 | 1 | US-08-152-039A-32    | Sequence 32, Appl  |
| 700 | 6 | 0.8 | 177 | 4 | US-09-344-837A-124   | Sequence 124, App  | 763 | 219 | 1 | US-08-557-309B-54    | Sequence 54, Appl  |
| 701 | 6 | 0.8 | 177 | 4 | US-09-344-837A-125   | Sequence 125, App  | 764 | 219 | 3 | US-08-834-776A-3     | Sequence 3, Appl   |
| 702 | 6 | 0.8 | 177 | 4 | US-09-344-837A-126   | Sequence 126, App  | 765 | 219 | 3 | US-09-527-345-2      | Sequence 2, Appl   |
| 703 | 6 | 0.8 | 177 | 4 | US-09-344-837A-127   | Sequence 127, App  | 766 | 219 | 4 | US-09-252-991A-19313 | Sequence 19313, A  |
| 704 | 6 | 0.8 | 177 | 4 | US-09-344-837A-128   | Sequence 128, App  | 767 | 219 | 4 | US-09-252-991A-23215 | Sequence 23215, A  |
| 705 | 6 | 0.8 | 177 | 4 | US-09-558-089-8      | Sequence 8, Appl   | 768 | 220 | 4 | US-09-252-991A-28993 | Sequence 28993, A  |
| 706 | 6 | 0.8 | 177 | 4 | US-09-558-089-8      | Sequence 8, Appl   | 769 | 220 | 4 | US-09-252-991A-18874 | Sequence 18874, A  |
| 707 | 6 | 0.8 | 177 | 4 | US-09-558-089-8      | Sequence 8, Appl   | 770 | 221 | 4 | US-09-252-991A-24616 | Sequence 24616, A  |
| 708 | 6 | 0.8 | 177 | 4 | US-09-558-089-8      | Sequence 8, Appl   | 771 | 221 | 4 | US-09-107-532A-6667  | Sequence 6667, Ap  |
| 709 | 6 | 0.8 | 177 | 4 | US-09-558-089-8      | Sequence 8, Appl   | 772 | 222 | 4 | US-09-252-991A-23841 | Sequence 23841, A  |
| 710 | 6 | 0.8 | 179 | 3 | US-09-252-991A-26853 | Sequence 2, Appl   | 773 | 222 | 4 | US-09-166-350-20     | Sequence 20, Appl  |
| 711 | 6 | 0.8 | 180 | 6 | 5514590-9            | Patent No. 5514590 | 774 | 223 | 4 | US-09-461-325-494    | Sequence 494, Appl |
| 712 | 6 | 0.8 | 181 | 1 | US-08-278-091-15     | Sequence 15, Appl  | 775 | 223 | 4 | US-09-252-991A-19065 | Sequence 19065, A  |
| 713 | 6 | 0.8 | 181 | 1 | US-08-483-859-15     | Sequence 15, Appl  | 776 | 224 | 4 | US-09-328-352-4696   | Sequence 4696, Ap  |
| 714 | 6 | 0.8 | 181 | 1 | US-08-472-173-15     | Sequence 15, Appl  | 777 | 225 | 4 | US-09-328-352-4696   | Sequence 17754, A  |
| 715 | 6 | 0.8 | 181 | 2 | US-08-487-167-15     | Sequence 15, Appl  | 778 | 226 | 3 | US-09-252-991A-17754 | Sequence 30, Appl  |
| 716 | 6 | 0.8 | 181 | 2 | US-08-482-816-15     | Sequence 15, Appl  | 779 | 227 | 4 | US-09-195-868-30     | Sequence 31689, A  |
| 717 | 6 | 0.8 | 181 | 2 | US-08-296-149-15     | Sequence 15, Appl  | 780 | 228 | 4 | US-09-252-991A-31689 | Sequence 28783, A  |
| 718 | 6 | 0.8 | 181 | 2 | US-08-801-499-15     | Sequence 15, Appl  | 781 | 229 | 4 | US-09-194-468A-23    | Sequence 21744, A  |
| 719 | 6 | 0.8 | 181 | 2 | US-08-615-271-15     | Sequence 15, Appl  | 782 | 230 | 4 | US-09-252-991A-21744 | Sequence 22207, A  |
| 720 | 6 | 0.8 | 181 | 3 | US-09-074-659-15     | Sequence 15, Appl  | 783 | 230 | 4 | US-09-252-991A-22207 | Sequence 22765, A  |
| 721 | 6 | 0.8 | 181 | 3 | US-09-106-468-15     | Sequence 15, Appl  | 784 | 232 | 3 | US-09-252-991A-22765 | Sequence 8, Appl   |
| 722 | 6 | 0.8 | 181 | 3 | US-09-106-468-15     | Sequence 15, Appl  | 785 | 233 | 4 | US-09-201-641-8      | Sequence 517, App  |
| 723 | 6 | 0.8 | 182 | 1 | US-09-252-991A-26910 | Sequence 26910, A  | 786 | 234 | 4 | US-08-936-165A-517   | Sequence 7, Appl   |
| 724 | 6 | 0.8 | 182 | 1 | US-07-948-812-6      | Sequence 6, Appl   | 787 | 234 | 4 | US-09-219-983A-7     | Sequence 353, App  |
| 725 | 6 | 0.8 | 182 | 4 | US-09-634-238-370    | Sequence 370, App  | 788 | 234 | 4 | US-09-149-476-353    | Sequence 28420, A  |
| 726 | 6 | 0.8 | 183 | 4 | US-09-143-211-16     | Sequence 18, Appl  | 789 | 235 | 3 | US-09-807-151-3      | Sequence 3, Appl   |
| 727 | 6 | 0.8 | 183 | 4 | US-09-252-991A-29126 | Sequence 23, Appl  | 790 | 235 | 4 | US-09-478-957-3      | Sequence 13, Appl  |
| 728 | 6 | 0.8 | 184 | 3 | US-08-149-101A-23    | Sequence 23, Appl  | 791 | 236 | 3 | US-09-049-672A-13    | Sequence 13, Appl  |
| 729 | 6 | 0.8 | 184 | 5 | PCT-US94-12873-23    | Sequence 23, Appl  | 792 | 237 | 1 | US-08-910-973-13     | Sequence 3, Appl   |
| 730 | 6 | 0.8 | 185 | 4 | US-09-252-991A-24804 | Sequence 24804, A  | 793 | 237 | 3 | US-08-818-514-3      | Sequence 3, Appl   |
| 731 | 6 | 0.8 | 185 | 4 | US-09-252-991A-24804 | Sequence 18, App   | 794 | 237 | 3 | US-09-115-934A-3     | Sequence 13, Appl  |
| 732 | 6 | 0.8 | 185 | 4 | US-09-198-452A-184   | Sequence 184, App  | 795 | 237 | 3 | US-09-499-227-13     | Sequence 7, Appl   |
| 733 | 6 | 0.8 | 187 | 4 | US-08-916-246-12     | Sequence 6, Appl   | 796 | 237 | 4 | US-09-499-227-13     | Sequence 3, Appl   |
| 734 | 6 | 0.8 | 187 | 4 | US-09-534-407-6      | Sequence 6, Appl   | 797 | 237 | 4 | US-09-611-175-3      | Sequence 7, Appl   |
| 735 | 6 | 0.8 | 187 | 4 | US-09-999-201B-6     | Sequence 6, Appl   | 798 | 239 | 4 | US-08-851-971-7      | Sequence 19895, A  |
| 736 | 6 | 0.8 | 187 | 4 | US-09-252-991A-29442 | Sequence 29442, A  | 799 | 239 | 4 | US-09-252-991A-18995 | Sequence 21250, A  |
| 737 | 6 | 0.8 | 188 | 4 | US-09-252-991A-20399 | Sequence 20399, A  | 800 | 239 | 4 | US-09-252-991A-21250 | Sequence 25790, A  |
| 738 | 6 | 0.8 | 190 | 4 | US-09-252-991A-17963 | Sequence 17963, A  | 801 | 239 | 4 | US-09-252-991A-25790 | Sequence 2, Appl   |
| 739 | 6 | 0.8 | 191 | 4 | US-09-252-991A-28753 | Sequence 28753, A  | 802 | 240 | 2 | US-07-956-399-2      | Sequence 29766, A  |
| 740 | 6 | 0.8 | 191 | 4 | US-09-252-991A-28753 | Sequence 28929, A  | 803 | 240 | 4 | US-09-252-991A-29766 | Sequence 4430, Ap  |
| 741 | 6 | 0.8 | 192 | 4 | US-09-252-991A-29829 | Sequence 23163, A  | 804 | 240 | 4 | US-09-107-532A-4430  | Sequence 46, Appl  |
| 742 | 6 | 0.8 | 193 | 4 | US-09-252-991A-23163 | Sequence 24, Appl  | 805 | 241 | 3 | US-08-961-083-46     | Sequence 46, Appl  |
| 743 | 6 | 0.8 | 194 | 4 | US-09-194-468A-44    | Sequence 20731, A  | 806 | 241 | 4 | US-09-536-784-46     | Sequence 4086, Ap  |
| 744 | 6 | 0.8 | 196 | 4 | US-09-752-165-95     | Sequence 95, Appl  | 807 | 241 | 4 | US-09-107-532A-4086  | Sequence 23, Appl  |
| 745 | 6 | 0.8 | 196 | 4 | US-09-252-991A-29477 | Sequence 29477, A  | 808 | 241 | 4 | US-08-220-602B-23    | Sequence 19853, A  |
| 746 | 6 | 0.8 | 196 | 4 | US-09-252-991A-30527 | Sequence 30527, A  | 809 | 242 | 4 | US-09-252-991A-28448 | Sequence 22, Appl  |
| 747 | 6 | 0.8 | 197 | 3 | US-09-370-473-12     | Sequence 12, Appl  | 810 | 242 | 4 | US-09-252-991A-19853 | Sequence 22, Appl  |
| 748 | 6 | 0.8 | 197 | 4 | US-09-252-991A-32518 | Sequence 32518, A  | 811 | 244 | 1 | US-08-910-973-22     | Sequence 30810, A  |
| 749 | 6 | 0.8 | 199 | 4 | US-08-936-165A-388   | Sequence 388, App  | 812 | 244 | 4 | US-09-499-227-22     | Sequence 1, Appl   |
| 750 | 6 | 0.8 | 200 | 4 | US-09-252-991A-44828 | Sequence 24828, A  | 813 | 244 | 4 | US-09-252-991A-30810 | Sequence 1, Appl   |
| 751 | 6 | 0.8 | 201 | 4 | US-09-252-991A-32202 | Sequence 32202, A  | 814 | 244 | 4 | US-08-722-570-1      | Sequence 14, Appl  |
| 752 | 6 | 0.8 | 204 | 1 | US-08-792-019B-10    | Sequence 10, Appl  | 815 | 244 | 4 | US-08-932-411A-1     | Sequence 8, Appl   |
| 753 | 6 | 0.8 | 204 | 3 | US-08-988-819-10     | Sequence 10, Appl  | 816 | 244 | 4 | US-08-932-411A-14    | Sequence 24933, A  |
| 754 | 6 | 0.8 | 204 | 3 | US-09-016-53A-10     | Sequence 5, Appl   | 817 | 244 | 1 | US-08-015-985-8      | Sequence 5, Appl   |
| 755 | 6 | 0.8 | 204 | 3 | US-08-097-669-5      | Sequence 8, Appl   | 818 | 245 | 4 | US-09-352-991A-24933 | Sequence 21906, A  |
| 756 | 6 | 0.8 | 205 | 4 | US-09-786-240-8      | Sequence 3, Appl   | 819 | 246 | 4 | US-09-352-991A-21908 | Sequence 23967, A  |
| 757 | 6 | 0.8 | 207 | 4 | US-09-143-211-3      | Sequence 23638, A  | 820 | 246 | 4 | US-09-352-991A-23967 | Sequence 5431, Ap  |
| 758 | 6 | 0.8 | 208 | 4 | US-09-252-991A-23638 | Sequence 23638, A  | 821 | 246 | 4 | US-09-328-352-5431   | Sequence 5045, Ap  |
| 759 | 6 | 0.8 | 208 | 4 | US-09-328-352-6813   | Sequence 6813, Ap  | 822 | 246 | 4 | US-09-328-352-5045   | Sequence 22391, A  |
| 760 | 6 | 0.8 | 209 | 4 | US-09-252-991A-16676 | Sequence 16676, A  | 823 | 247 | 4 | US-09-252-991A-21188 | Sequence 29850, A  |
| 761 | 6 | 0.8 | 209 | 4 | US-09-252-991A-16676 | Sequence 5469, Ap  | 824 | 248 | 4 | US-09-252-991A-29850 | Sequence 2, Appl   |
| 762 | 6 | 0.8 | 212 | 3 | US-09-041-889-1      | Sequence 1, Appl   | 825 | 250 | 1 | US-08-562-311-2      | Sequence 10, Appl  |
| 763 | 6 | 0.8 | 212 | 3 | US-08-837-058-1      | Sequence 1, Appl   | 826 | 250 | 3 | US-09-656-450-10     | Sequence 10, Appl  |
| 764 | 6 | 0.8 | 212 | 3 | US-09-996-243-387    | Sequence 387, App  | 827 | 250 | 3 |                      |                    |
| 765 | 6 | 0.8 | 213 | 4 | US-09-417-264-1      | Sequence 1, Appl   | 828 | 250 | 3 |                      |                    |
| 766 | 6 | 0.8 | 215 | 4 | US-09-252-991A-21254 | Sequence 21254, A  | 829 | 250 | 3 |                      |                    |
| 767 | 6 | 0.8 | 215 | 4 | US-09-252-991A-20807 | Sequence 20807, A  | 830 | 250 | 3 |                      |                    |
| 768 | 6 | 0.8 | 216 | 4 | US-09-252-991A-26828 | Sequence 26828, A  |     |     |   |                      |                    |

|     |   |     |     |   |                      |                      |     |   |     |     |   |                      |                    |
|-----|---|-----|-----|---|----------------------|----------------------|-----|---|-----|-----|---|----------------------|--------------------|
| 831 | 6 | 0.8 | 251 | 4 | US-09-252-991A-28124 | Sequence 28124, A    | 904 | 6 | 0.8 | 299 | 4 | US-09-252-991A-18498 | Sequence 18498, A  |
| 832 | 6 | 0.8 | 252 | 4 | US-09-395-689-5      | Sequence 5, Appli    | 905 | 6 | 0.8 | 300 | 4 | US-09-585-858-41     | Sequence 41, Appli |
| 833 | 6 | 0.8 | 253 | 4 | US-09-134-001C-4565  | Sequence 4565, Ap    | 906 | 6 | 0.8 | 301 | 2 | US-08-355-844-2      | Sequence 2, Appli  |
| 834 | 6 | 0.8 | 254 | 4 | US-09-252-991A-32425 | Sequence 32425, A    | 907 | 6 | 0.8 | 301 | 3 | US-08-303-861-21     | Sequence 21, Appli |
| 835 | 6 | 0.8 | 255 | 4 | US-09-252-991A-17342 | Sequence 17342, A    | 908 | 6 | 0.8 | 301 | 3 | US-09-011-073A-1     | Sequence 1, Appli  |
| 836 | 6 | 0.8 | 256 | 4 | US-09-107-532A-4208  | Sequence 4208, Ap    | 909 | 6 | 0.8 | 301 | 3 | US-09-230-421-2      | Sequence 2, Appli  |
| 837 | 6 | 0.8 | 257 | 4 | US-09-266-965-102    | Sequence 102, App    | 910 | 6 | 0.8 | 301 | 4 | US-09-347-504-12     | Sequence 12, Appli |
| 838 | 6 | 0.8 | 258 | 4 | US-09-252-991A-29749 | Sequence 29749, A    | 911 | 6 | 0.8 | 301 | 4 | US-08-902-572-14     | Sequence 14, Appli |
| 839 | 6 | 0.8 | 259 | 4 | US-09-648-004-4      | Sequence 4, Appli    | 912 | 6 | 0.8 | 301 | 4 | US-09-773-430-1      | Sequence 1, Appli  |
| 840 | 6 | 0.8 | 260 | 4 | US-09-134-001C-4872  | Sequence 4872, Ap    | 913 | 6 | 0.8 | 301 | 5 | PCT-US95-16126-2     | Sequence 2, Appli  |
| 841 | 6 | 0.8 | 261 | 3 | US-08-768-373-2      | Sequence 2, Appli    | 914 | 6 | 0.8 | 302 | 4 | US-09-252-991A-24353 | Sequence 24353, A  |
| 842 | 6 | 0.8 | 262 | 4 | US-09-328-352-8087   | Sequence 8087, Ap    | 915 | 6 | 0.8 | 304 | 1 | US-08-132-405-4      | Sequence 4, Appli  |
| 843 | 6 | 0.8 | 263 | 1 | US-08-106-981-6      | Sequence 6, Appli    | 916 | 6 | 0.8 | 304 | 1 | US-08-395-939A-4     | Sequence 4, Appli  |
| 844 | 6 | 0.8 | 264 | 1 | US-08-403-379A-1     | Sequence 1, Appli    | 917 | 6 | 0.8 | 304 | 5 | PCT-US91-01861-4     | Sequence 4, Appli  |
| 845 | 6 | 0.8 | 265 | 2 | US-08-929-414-1      | Sequence 1, Appli    | 918 | 6 | 0.8 | 305 | 4 | US-09-634-238-320    | Sequence 320, App  |
| 846 | 6 | 0.8 | 266 | 2 | US-08-695-992-5      | Sequence 5, Appli    | 919 | 6 | 0.8 | 305 | 4 | US-09-252-991A-22922 | Sequence 22922, A  |
| 847 | 6 | 0.8 | 267 | 2 | US-08-557-309B-51    | Sequence 51, Appli   | 920 | 6 | 0.8 | 306 | 1 | US-08-279-754-2      | Sequence 2, Appli  |
| 848 | 6 | 0.8 | 268 | 3 | US-09-144-525-5      | Sequence 5, Appli    | 921 | 6 | 0.8 | 306 | 4 | US-09-252-991A-18873 | Sequence 18873, A  |
| 849 | 6 | 0.8 | 269 | 3 | US-08-834-306-51     | Sequence 51, Appli   | 922 | 6 | 0.8 | 306 | 5 | PCT-US95-09052-2     | Sequence 2, Appli  |
| 850 | 6 | 0.8 | 270 | 3 | US-08-993-674A-51    | Sequence 51, Appli   | 923 | 6 | 0.8 | 307 | 3 | US-08-469-318-121    | Sequence 121, App  |
| 851 | 6 | 0.8 | 271 | 4 | US-09-134-001C-4017  | Sequence 4017, Ap    | 924 | 6 | 0.8 | 307 | 3 | US-08-469-318-122    | Sequence 122, App  |
| 852 | 6 | 0.8 | 272 | 4 | US-09-256-976-51     | Sequence 51, Appli   | 925 | 6 | 0.8 | 307 | 3 | US-08-469-318-123    | Sequence 123, App  |
| 853 | 6 | 0.8 | 273 | 4 | US-09-107-532A-6468  | Sequence 6468, Ap    | 926 | 6 | 0.8 | 307 | 3 | US-08-469-318-124    | Sequence 124, App  |
| 854 | 6 | 0.8 | 274 | 3 | US-09-053-197A-8     | Sequence 8, Appli    | 927 | 6 | 0.8 | 307 | 3 | US-08-469-318-134    | Sequence 134, App  |
| 855 | 6 | 0.8 | 275 | 4 | US-09-085-761A-8     | Sequence 8, Appli    | 928 | 6 | 0.8 | 307 | 3 | US-08-469-318-135    | Sequence 135, App  |
| 856 | 6 | 0.8 | 276 | 4 | US-09-198-452A-715   | Sequence 715, App    | 929 | 6 | 0.8 | 307 | 3 | US-08-469-318-146    | Sequence 146, App  |
| 857 | 6 | 0.8 | 277 | 4 | US-09-252-991A-32512 | Sequence 32512, A    | 930 | 6 | 0.8 | 307 | 3 | US-08-469-318-147    | Sequence 147, App  |
| 858 | 6 | 0.8 | 278 | 4 | US-09-252-991A-28620 | Sequence 28620, A    | 931 | 6 | 0.8 | 307 | 3 | US-08-469-318-152    | Sequence 152, App  |
| 859 | 6 | 0.8 | 279 | 4 | US-09-252-991A-21630 | Sequence 21630, A    | 932 | 6 | 0.8 | 307 | 3 | US-08-469-318-158    | Sequence 158, App  |
| 860 | 6 | 0.8 | 280 | 4 | US-09-252-991A-24063 | Sequence 24063, A    | 933 | 6 | 0.8 | 307 | 3 | US-08-469-318-159    | Sequence 159, App  |
| 861 | 6 | 0.8 | 281 | 4 | US-09-252-991A-17589 | Sequence 17589, A    | 934 | 6 | 0.8 | 307 | 3 | US-08-468-609A-121   | Sequence 121, App  |
| 862 | 6 | 0.8 | 282 | 2 | US-08-467-265-2      | Sequence 2, Appli    | 935 | 6 | 0.8 | 307 | 3 | US-08-468-609A-122   | Sequence 122, App  |
| 863 | 6 | 0.8 | 283 | 3 | US-08-467-265-2      | Sequence 2, Appli    | 936 | 6 | 0.8 | 307 | 3 | US-08-468-609A-123   | Sequence 123, App  |
| 864 | 6 | 0.8 | 284 | 3 | US-09-407-891-2      | Sequence 2, Appli    | 937 | 6 | 0.8 | 307 | 3 | US-08-468-609A-124   | Sequence 124, App  |
| 865 | 6 | 0.8 | 285 | 3 | US-09-198-452A-353   | Sequence 353, App    | 938 | 6 | 0.8 | 307 | 3 | US-08-468-609A-134   | Sequence 134, App  |
| 866 | 6 | 0.8 | 286 | 4 | US-09-252-991A-18313 | Sequence 18313, A    | 939 | 6 | 0.8 | 307 | 3 | US-08-468-609A-135   | Sequence 135, App  |
| 867 | 6 | 0.8 | 287 | 4 | US-09-252-991A-22707 | Sequence 22707, A    | 940 | 6 | 0.8 | 307 | 3 | US-08-468-609A-146   | Sequence 146, App  |
| 868 | 6 | 0.8 | 288 | 4 | US-09-252-991A-25432 | Sequence 25432, A    | 941 | 6 | 0.8 | 307 | 3 | US-08-468-609A-147   | Sequence 147, App  |
| 869 | 6 | 0.8 | 289 | 2 | US-09-987-080-75     | Sequence 75, Appli   | 942 | 6 | 0.8 | 307 | 3 | US-08-468-609A-152   | Sequence 152, App  |
| 870 | 6 | 0.8 | 290 | 2 | US-08-997-362-75     | Sequence 75, Appli   | 943 | 6 | 0.8 | 307 | 3 | US-08-468-609A-158   | Sequence 158, App  |
| 871 | 6 | 0.8 | 291 | 3 | US-08-997-362-75     | Sequence 75, Appli   | 944 | 6 | 0.8 | 307 | 3 | US-08-468-609A-159   | Sequence 159, App  |
| 872 | 6 | 0.8 | 292 | 3 | US-09-095-855-75     | Sequence 75, Appli   | 945 | 6 | 0.8 | 307 | 4 | US-08-446-872A-121   | Sequence 121, App  |
| 873 | 6 | 0.8 | 293 | 4 | US-09-334-542-75     | Sequence 75, Appli   | 946 | 6 | 0.8 | 307 | 4 | US-08-446-872A-122   | Sequence 122, App  |
| 874 | 6 | 0.8 | 294 | 4 | US-09-205-426-75     | Sequence 75, Appli   | 947 | 6 | 0.8 | 307 | 4 | US-08-446-872A-123   | Sequence 123, App  |
| 875 | 6 | 0.8 | 295 | 4 | US-09-252-991A-24694 | Sequence 24694, A    | 948 | 6 | 0.8 | 307 | 4 | US-08-446-872A-124   | Sequence 124, App  |
| 876 | 6 | 0.8 | 296 | 4 | US-09-198-452A-1104  | Sequence 1104, Ap    | 949 | 6 | 0.8 | 307 | 4 | US-08-446-872A-134   | Sequence 134, App  |
| 877 | 6 | 0.8 | 297 | 4 | US-09-328-352-7814   | Sequence 7814, Ap    | 950 | 6 | 0.8 | 307 | 4 | US-08-446-872A-135   | Sequence 135, App  |
| 878 | 6 | 0.8 | 298 | 4 | US-09-134-001C-3651  | Sequence 3651, Ap    | 951 | 6 | 0.8 | 307 | 4 | US-08-446-872A-146   | Sequence 146, App  |
| 879 | 6 | 0.8 | 299 | 4 | US-09-252-991A-21038 | Sequence 21038, A    | 952 | 6 | 0.8 | 307 | 4 | US-08-446-872A-152   | Sequence 152, App  |
| 880 | 6 | 0.8 | 300 | 4 | US-09-252-991A-30528 | Sequence 30528, A    | 953 | 6 | 0.8 | 307 | 4 | US-08-446-872A-158   | Sequence 158, App  |
| 881 | 6 | 0.8 | 301 | 4 | US-09-252-991A-16673 | Sequence 16673, A    | 954 | 6 | 0.8 | 307 | 4 | US-08-446-872A-159   | Sequence 159, App  |
| 882 | 6 | 0.8 | 302 | 4 | US-09-404-879A-393   | Sequence 393, App    | 955 | 6 | 0.8 | 307 | 4 | US-08-762-227A-121   | Sequence 121, App  |
| 883 | 6 | 0.8 | 303 | 4 | US-09-966-243-291    | Sequence 231, App    | 956 | 6 | 0.8 | 307 | 4 | US-08-762-227A-122   | Sequence 122, App  |
| 884 | 6 | 0.8 | 304 | 3 | US-09-135-021-6      | Sequence 6, Appli    | 957 | 6 | 0.8 | 307 | 4 | US-08-762-227A-123   | Sequence 123, App  |
| 885 | 6 | 0.8 | 305 | 3 | US-09-793-229-31     | Sequence 31, Appli   | 958 | 6 | 0.8 | 307 | 4 | US-08-762-227A-124   | Sequence 124, App  |
| 886 | 6 | 0.8 | 306 | 3 | US-09-285-957-31     | Sequence 31, Appli   | 959 | 6 | 0.8 | 307 | 4 | US-08-762-227A-134   | Sequence 134, App  |
| 887 | 6 | 0.8 | 307 | 3 | US-09-338-352-6357   | Sequence 6357, Ap    | 960 | 6 | 0.8 | 307 | 4 | US-08-762-227A-135   | Sequence 135, App  |
| 888 | 6 | 0.8 | 308 | 4 | US-09-252-991A-27746 | Sequence 27746, A    | 961 | 6 | 0.8 | 307 | 4 | US-08-762-227A-146   | Sequence 146, App  |
| 889 | 6 | 0.8 | 309 | 4 | US-09-252-991A-32274 | Sequence 32274, A    | 962 | 6 | 0.8 | 307 | 4 | US-08-762-227A-147   | Sequence 147, App  |
| 890 | 6 | 0.8 | 310 | 1 | US-08-386-650-1      | Sequence 1, Appli    | 963 | 6 | 0.8 | 307 | 4 | US-08-762-227A-152   | Sequence 152, App  |
| 891 | 6 | 0.8 | 311 | 1 | US-08-768-626-1      | Sequence 1, Appli    | 964 | 6 | 0.8 | 307 | 4 | US-08-762-227A-158   | Sequence 158, App  |
| 892 | 6 | 0.8 | 312 | 4 | US-09-134-001C-3333  | Sequence 3333, Ap    | 965 | 6 | 0.8 | 307 | 4 | US-08-762-227A-159   | Sequence 159, App  |
| 893 | 6 | 0.8 | 313 | 4 | US-09-107-532A-6483  | Sequence 6483, Ap    | 966 | 6 | 0.8 | 307 | 4 | US-09-252-991A-21588 | Sequence 21588, A  |
| 894 | 6 | 0.8 | 314 | 4 | US-09-167-206-2      | Sequence 22, Appli   | 967 | 6 | 0.8 | 307 | 4 | US-09-252-991A-24199 | Sequence 24199, A  |
| 895 | 6 | 0.8 | 315 | 4 | US-09-252-991A-20538 | Sequence 20538, A    | 968 | 6 | 0.8 | 307 | 4 | US-09-107-532A-6125  | Sequence 6125, Ap  |
| 896 | 6 | 0.8 | 316 | 4 | US-09-252-991A-31376 | Sequence 31376, A    | 969 | 6 | 0.8 | 307 | 5 | PCT-US95-01185-121   | Sequence 121, App  |
| 897 | 6 | 0.8 | 317 | 3 | US-09-036-482B-2     | Sequence 2, Appli    | 970 | 6 | 0.8 | 307 | 5 | PCT-US95-01185-122   | Sequence 122, App  |
| 898 | 6 | 0.8 | 318 | 3 | US-09-198-452A-268   | Sequence 268, App    | 971 | 6 | 0.8 | 307 | 5 | PCT-US95-01185-123   | Sequence 123, App  |
| 899 | 6 | 0.8 | 319 | 3 | US-09-198-452A-268   | Sequence 1, Appli    | 972 | 6 | 0.8 | 307 | 5 | PCT-US95-01185-124   | Sequence 124, App  |
| 900 | 6 | 0.8 | 320 | 3 | US-09-173-581-1      | Sequence 1, Appli    | 973 | 6 | 0.8 | 307 | 5 | PCT-US95-01185-134   | Sequence 134, App  |
| 901 | 6 | 0.8 | 321 | 3 | US-09-420-915-1      | Patent No. 5514590   | 974 | 6 | 0.8 | 307 | 5 | PCT-US95-01185-135   | Sequence 135, App  |
| 902 | 6 | 0.8 | 322 | 3 | US-09-420-915-1      | Patent No. 5514590-2 | 975 | 6 | 0.8 | 307 | 5 | PCT-US95-01185-146   | Sequence 146, App  |
| 903 | 6 | 0.8 | 323 | 4 | US-09-648-004-22     | Sequence 22, Appli   | 976 | 6 | 0.8 | 307 | 5 |                      |                    |



```
977 6 0.8 307 5 PCT-US95-01185-147 Sequence 147, App
978 6 0.8 307 5 PCT-US95-01185-152 Sequence 152, App
979 6 0.8 307 5 PCT-US95-01185-158 Sequence 158, App
980 6 0.8 307 5 PCT-US95-01185-159 Sequence 159, App
981 6 0.8 308 4 US-09-804-778A-2 Sequence 2, Appli
982 6 0.8 308 4 US-09-804-778A-4 Sequence 4, Appli
983 6 0.8 308 4 US-09-328-352-5565 Sequence 565, Ap
984 6 0.8 309 4 US-09-134-001C-3039 Sequence 3039, Ap
985 6 0.8 309 4 US-09-404-879A-392 Sequence 392, App
986 6 0.8 310 1 US-08-363-255-6 Sequence 6, Appli
987 6 0.8 310 4 US-09-328-352-5923 Sequence 5923, Ap
988 6 0.8 312 3 US-09-198-930-142 Sequence 142, App
989 6 0.8 312 4 US-09-312-283C-142 Sequence 142, App
990 6 0.8 314 4 US-09-544-618-20 Sequence 20, Appli
991 6 0.8 315 4 US-09-252-991A-20012 Sequence 20012, A
992 6 0.8 315 4 US-09-252-991A-26188 Sequence 26188, A
993 6 0.8 316 4 US-09-252-991A-32957 Sequence 32957, A
994 6 0.8 318 4 US-09-252-991A-23557 Sequence 23557, A
995 6 0.8 318 4 US-09-252-991A-31947 Sequence 31947, A
996 6 0.8 319 4 US-09-252-991A-23827 Sequence 23827, A
997 6 0.8 321 4 US-09-252-991A-29819 Sequence 29819, A
998 6 0.8 321 4 US-09-252-991A-29844 Sequence 29844, A
999 6 0.8 321 4 US-09-107-532A-4129 Sequence 4129, Ap
1000 6 0.8 322 2 US-08-428-414A-2 Sequence 2, Appli
```

ALIGNMENTS

```
RESULT 1
US-09-328-352-6048
; Sequence 6048, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6048
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6048
Query Match 1.4%; Score 11; DB 4; Length 739;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 386 LKDVWDQALD 396
Db 320 LKDVWDQALD 330
|||||
```

```
RESULT 2
US-09-252-991A-33073
; Sequence 33073, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 33073
```

```
; LENGTH: 751
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-33073
Query Match 1.4%; Score 11; DB 4; Length 751;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 515 QVMIEARIVEA 525
Db 486 QVMIEARIVEA 496
|||||
```

```
RESULT 3
US-08-911-853-15
; Sequence 15, Application US/08911853
; Patent No. 6048710
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gijlsbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,853
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,092
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 649 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-911-853-15
```

```
Query Match 1.2%; Score 9; DB 3; Length 649;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 725 KVPLLGDIP 733
Db 563 KVPLLGDIP 571
|||||
```

```
RESULT 4
US-09-479-409-15
; Sequence 15, Application US/09479409
; Patent No. 6225106
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gijlsbert
; APPLICANT: Quax, Wilhelmus J.
```

```

; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/479,409
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/911,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 649 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-479-409-15
;
; Query Match 1.2%; Score 9; DB 3; Length 649;
; Best Local Similarity 100.0%; Pred. No. 7.6;
; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 725 KVPLLGDIP 733
; Db 563 KVPLLGDIP 571
;
; RESULT 5
; US-09-479-453-15
; Sequence 15, Application US/09479453
; Patent No. 6313283
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gijbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/479,453
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/911,853
; FILING DATE:

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 649 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-479-453-15
;
; Query Match 1.2%; Score 9; DB 4; Length 649;
; Best Local Similarity 100.0%; Pred. No. 7.6;
; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 725 KVPLLGDIP 733
; Db 563 KVPLLGDIP 571
;
; RESULT 6
; US-09-252-991A-28934
; Sequence 28934, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28934
; LENGTH: 753
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-28934

```

```

; Query Match 1.2%; Score 9; DB 4; Length 753;
; Best Local Similarity 100.0%; Pred. No. 8.7;
; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 725 KVPLLGDIP 733
; Db 664 KVPLLGDIP 672
;
; RESULT 7
; US-09-252-991A-30225
; Sequence 30225, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30225
; LENGTH: 828

```

```

; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30225

Query Match      1.0%; Score 8; DB 4; Length 828;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      726 VPLGGDIP 733
DB      614 VPLGGDIP 621

RESULT 8
US-09-562-737-49
; Sequence 49, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 49
; LENGTH: 1024
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-09-562-737-49

Query Match      1.0%; Score 8; DB 4; Length 1024;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      194 QAAASAXQ 201
DB      923 QAAASAXQ 930

RESULT 9
US-09-187-859-2429
; Sequence 2429, Application US/09187859A
; Patent No. 6358920
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
; FILE REFERENCE: 100086.407C1
; CURRENT APPLICATION NUMBER: US/09/187,859A
; CURRENT FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 4052
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2429
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Representative cyclic modulating agent based on
; OTHER INFORMATION: PB-cadherin cell adhesion recognition sequence
US-09-187-859-2429

Query Match      0.9%; Score 7; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      504 KLIDELD 510
DB      1 KLIDELD 7

RESULT 10
US-09-839-542B-2429
; Sequence 2429, Application US/09839542B
; Patent No. 6569996
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Symonds, James Matthew
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
; FILE REFERENCE: 100086.407D1
; CURRENT APPLICATION NUMBER: US/09/839,542B
; CURRENT FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 4052
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2429
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Representative cyclic modulating agent based on
; OTHER INFORMATION: PB-cadherin cell adhesion recognition sequence
US-09-839-542B-2429

Query Match      0.9%; Score 7; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      504 KLIDELD 510
DB      1 KLIDELD 7

RESULT 11
US-08-716-317-32
; Sequence 32, Application US/08716317
; Patent No. 5919654
; GENERAL INFORMATION:
; APPLICANT: HAMA, YUKO
; APPLICANT: TOMIDA, HIDEKI
; APPLICANT: TSUKAMOTO, HIROKO
; APPLICANT: NIKAIIDO, KIYOKAZU
; APPLICANT: KUMAGAI, HIROMICHI
; TITLE OF INVENTION: SECRETION SIGNAL GENE AND EXPRESSION
; TITLE OF INVENTION: VECTOR CONTAINING IT
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716,317
; FILING DATE: 02-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/00198
; FILING DATE: 01-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 17167/1995
; FILING DATE: 03-FEB-1996
; ATTORNEY/AGENT INFORMATION:

```

NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 59-924-0 PCT  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Peptide  
US-08-716-317-32

Query Match 0.9%; Score 7; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 181 VVVSAP 187  
Db 4 VVVSAP 10

RESULT 12  
US-08-716-317-33  
Sequence 33, Application US/08716317  
Patent No. 5919654  
GENERAL INFORMATION:  
APPLICANT: HAMA, YUKO  
APPLICANT: TOHDA, HIDEKI  
APPLICANT: TSUKAMOTO, HIROKO  
APPLICANT: NIKAI, KAZUO  
APPLICANT: KIMURA, HIROMICHI  
TITLE OF INVENTION: SECRETION SIGNAL GENE AND EXPRESSION  
TITLE OF INVENTION: VECTOR CONTAINING IT  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 02-OCT-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP96/00198  
FILING DATE: 01-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 17167/1995  
FILING DATE: 03-FEB-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 59-924-0 PCT  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
US-08-716-317-33

Query Match 0.9%; Score 7; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 181 VVVSAP 187  
Db 2 VVVSAP 8

RESULT 13  
US-08-107-676-7  
Sequence 7, Application US/08107676  
Patent No. 5955356  
GENERAL INFORMATION:  
APPLICANT: Content, Jean  
APPLICANT: Dewit, Lucas  
APPLICANT: DeBruyn, Jacqueline  
TITLE OF INVENTION: Mycobacterium Polypeptides and Nucleic  
TITLE OF INVENTION: Acids Encoding Them for Diagnosis and Control of  
TITLE OF INVENTION: Tuberculosis  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant & Gould  
STREET: 3100 No. 5955356west Center  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/107,676  
FILING DATE: 04-OCT-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/EP92/00268  
FILING DATE: 07-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 91400388.4  
FILING DATE: 14-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kowalchuk, Katherine M.  
REGISTRATION NUMBER: 36,848  
REFERENCE/DOCKET NUMBER: 8076.89USWO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-332-5300  
TELEFAX: 612-332-9081  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-107-676-7

Query Match 0.9%; Score 7; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 141 AAPA 147  
Db 3 AAPA 9

RESULT 14  
US-08-107-676-22  
Sequence 22, Application US/08107676

Patent No. 5955356  
; GENERAL INFORMATION:  
; APPLICANT: Content, Jean  
; APPLICANT: DeMit, Lucas  
; APPLICANT: DeBruyn, Jacqueline  
; TITLE OF INVENTION: Mycobacterium Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding Them for Diagnosis and Control of  
; TITLE OF INVENTION: Tuberculosis  
; NUMBER OF SEQUENCES: 38  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 3100 No. 5955356west Center  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/107,676  
; FILING DATE: 04-OCT-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/EP92/00268  
; FILING DATE: 07-FEB-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 91400388.4  
; FILING DATE: 14-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kowalchuk, Katherine M.  
; REGISTRATION NUMBER: 36,848  
; REFERENCE/DOCKET NUMBER: 8076.89USWO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-332-5300  
; TELEFAX: 612-332-9081  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-107-676-22  
Query Match 0.9%; Score 7; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 141 AAPAAPA 147  
Db 13 AAPAAPA 19  
RESULT 15  
US-09-194-296-36  
; Sequence 36, Application US/09194296  
; Patent No. 6486125  
; GENERAL INFORMATION:  
; APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA  
; TITLE OF INVENTION: SYNTHETIC APPROACH TO DESIGNED CHEMICAL  
; STRUCTURES  
; NUMBER OF SEQUENCES: 38  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MUETING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.  
; STREET: 119 No. 6486125th Fourth Street, Suite 203  
; CITY: Minneapolis  
; STATE: Minnesota  
; COUNTRY: U.S.A.  
; ZIP: 55401  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/194,296  
FILING DATE: 15-Oct-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/653,632  
FILING DATE: 24-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: MCCORMACK, MYRA M.  
REGISTRATION NUMBER: 36,602  
REFERENCE/DOCKET NUMBER: 110.00330220  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-305-1225  
TELEFAX: 612-305-1228  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 36:  
US-09-194-296-36  
Query Match 0.9%; Score 7; DB 4; Length 33;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 345 GRKISLD 351  
Db 27 GRKISLD 33  
RESULT 16  
US-08-499-676A-23  
; Sequence 23, Application US/08499676A  
; Patent No. 5998154  
; GENERAL INFORMATION:  
; APPLICANT: AGNES SCHONBRUNN  
; TITLE OF INVENTION: SOMATOSTATIN RECEPTOR PEPTIDE  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pravel, Hewitt, Kimball & Krieger  
; STREET: 1177 West Loop South, 10th Floor  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77027-9095  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/499,676A  
; FILING DATE: July 7, 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Krieger, Paul E.  
; REGISTRATION NUMBER: 25,886  
; REFERENCE/DOCKET NUMBER: 79247.3/A95175US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 713-850-0909  
; TELEFAX: 713-850-0165  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 42 amino acids

; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 US-08-499-676A-23

Query Match 0.9%; Score 7; DB 2; Length 42;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 NLESGGV 336  
 |||||  
 Db 23 NLESGGV 29

## RESULT 17

US-08-499-676A-24  
 ; Sequence 24, Application US/08499676A  
 ; Patent No. 5998154  
 ; GENERAL INFORMATION:  
 ; APPLICANT: AGNES SCHONBRUNN  
 ; TITLE OF INVENTION: SOMATOSTATIN RECEPTOR PEPTIDE  
 ; NUMBER OF SEQUENCES: 40  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pravel, Hewitt, Kimball & Krieger  
 ; STREET: 1177 West Loop South, 10th Floor  
 ; CITY: Houston  
 ; STATE: TX  
 ; COUNTRY: USA  
 ; ZIP: 77027-9095  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/499,676A  
 ; FILING DATE: July 7, 1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Krieger, Paul E.  
 ; REGISTRATION NUMBER: 25,885  
 ; REFERENCE/DOCKET NUMBER: 79247.3/A95175US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 713-850-0909  
 ; TELEFAX: 713-850-0165  
 ; INFORMATION FOR SEQ ID NO: 24:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 42 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 US-08-499-676A-24

Query Match 0.9%; Score 7; DB 2; Length 42;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 NLESGGV 336  
 |||||  
 Db 23 NLESGGV 29

## RESULT 18

US-08-905-223-472  
 ; Sequence 472, Application US/08905223  
 ; Patent No. 6222029  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Edwards, Jean-Baptiste D.

; APPLICANT: Duelt, Aymeric  
 ; APPLICANT: Lacroix, Bruno  
 ; TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS  
 ; NUMBER OF SEQUENCES: 503  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear  
 ; STREET: 501 West Broadway  
 ; CITY: San Diego  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 92101-3505  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy Disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: Win95  
 ; SOFTWARE: Word  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/905,223  
 ; FILING DATE:  
 ; CLASSIFICATION: 536  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Israelset, Ned A.  
 ; REGISTRATION NUMBER: 29,655  
 ; REFERENCE/DOCKET NUMBER:  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (619) 235-8550  
 ; TELEFAX: (619) 235-0176  
 ; INFORMATION FOR SEQ ID NO: 472:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 50 amino acids  
 ; TYPE: AMINO ACID  
 ; TOPOLOGY: LINEAR  
 ; MOLECULE TYPE: PROTEIN  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Homo Sapiens  
 ; TISSUE TYPE: Brain  
 ; FEATURE:  
 ; NAME/KEY: sig\_peptide  
 ; LOCATION: 16...1  
 ; IDENTIFICATION METHOD: Von Heijne matrix  
 ; OTHER INFORMATION: score 4.5  
 ; OTHER INFORMATION: seq RLLRRFLASVIS/RK  
 US-08-905-223-472

Query Match 0.9%; Score 7; DB 3; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 655 GLTVTPN 661  
 |||||  
 Db 41 GLTVTPN 47

## RESULT 19

US-08-499-676A-1  
 ; Sequence 1, Application US/08499676A  
 ; Patent No. 5998154  
 ; GENERAL INFORMATION:  
 ; APPLICANT: AGNES SCHONBRUNN  
 ; TITLE OF INVENTION: SOMATOSTATIN RECEPTOR PEPTIDE  
 ; NUMBER OF SEQUENCES: 40  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pravel, Hewitt, Kimball & Krieger  
 ; STREET: 1177 West Loop South, 10th Floor  
 ; CITY: Houston  
 ; STATE: TX  
 ; COUNTRY: USA  
 ; ZIP: 77027-9095  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/499,676A  
;; FILING DATE: July 7, 1995  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER:  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Krieger, Paul E.  
;; REGISTRATION NUMBER: 25,886  
;; REFERENCE/DOCKET NUMBER: 79247.3/A95175US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 713-850-0909  
;; TELEFAX: 713-850-0165  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 70 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
US-08-499-676A-1  
Query Match 0.9%; Score 7; DB 2; Length 70;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 330 NLESGGV 336  
Db 51 NLESGGV 57  
RESULT 20  
US-08-499-676A-2  
; Sequence 2, Application US/08499676A  
; Patent No. 5998154  
; GENERAL INFORMATION:  
; APPLICANT: AGNES SCHONBRUNN  
; TITLE OF INVENTION: SOMATOSTATIN RECEPTOR PEPTIDE  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pravel, Hewitt, Kimball & Krieger  
; STREET: 1177 West Loop South, 10th Floor  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77027-9095  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/499,676A  
; FILING DATE: July 7, 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Krieger, Paul E.  
; REGISTRATION NUMBER: 25,886  
; REFERENCE/DOCKET NUMBER: 79247.3/A95175US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 713-850-0909  
; TELEFAX: 713-850-0165  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 70 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide

US-08-499-676A-2  
Query Match 0.9%; Score 7; DB 2; Length 70;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 330 NLESGGV 336  
Db 51 NLESGGV 57  
RESULT 21  
US-08-499-676A-3  
; Sequence 3, Application US/08499676A  
; Patent No. 5998154  
; GENERAL INFORMATION:  
; APPLICANT: AGNES SCHONBRUNN  
; TITLE OF INVENTION: SOMATOSTATIN RECEPTOR PEPTIDE  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pravel, Hewitt, Kimball & Krieger  
; STREET: 1177 West Loop South, 10th Floor  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77027-9095  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/499,676A  
; FILING DATE: July 7, 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Krieger, Paul E.  
; REGISTRATION NUMBER: 25,886  
; REFERENCE/DOCKET NUMBER: 79247.3/A95175US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 713-850-0909  
; TELEFAX: 713-850-0165  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 70 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-499-676A-3  
Query Match 0.9%; Score 7; DB 2; Length 70;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 330 NLESGGV 336  
Db 51 NLESGGV 57  
RESULT 22  
US-08-252-991A-25964  
; Sequence 25964, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A

;; CURRENT FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: US 60/074,788  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 25964  
;; LENGTH: 79  
;; TYPE: PRT  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25964

Query Match 0.9%; Score 7; DB 4; Length 79;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPA 147  
Db 30 AAPAAPA 36

RESULT 23  
US-09-252-991A-23927  
; Sequence 23927, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23927  
; LENGTH: 85  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23927

Query Match 0.9%; Score 7; DB 4; Length 85;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 SAPARPA 138  
Db 71 SAPARPA 77

RESULT 24  
US-09-252-991A-21928  
; Sequence 21928, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21928  
; LENGTH: 98  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21928

Query Match 0.9%; Score 7; DB 4; Length 98;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPA 147  
Db 51 AAPAAPA 57

RESULT 25  
US-07-754-918A-12  
; Sequence 12, Application US/07754918A  
; Patent No. 5286484  
; GENERAL INFORMATION:  
; APPLICANT: Rodriguez, R.S. et al  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE CODING FOR AN  
; TITLE OF INVENTION: OUTER MEMBRANE PROTEIN FROM NEISSERIA MENINGITIDIS AND USE  
; TITLE OF INVENTION: OF SAID PROTEIN IN VACCINE PREPARATIONS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Stanger, Michaelson, Spivak and Wallace, Esq.  
; STREET: Parkway 109 Office Center, 328 Newman Springs  
; CITY: Red Bank  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5 1/4" 360Kb IBM compatible diskette  
; COMPUTER: IBM PS/2 Model 80  
; OPERATING SYSTEM: MS-DOS 5.0  
; SOFTWARE: Microsoft Word 5.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/754,918A  
; FILING DATE: 19910905  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michaelson, Peter L.  
; REGISTRATION NUMBER: 30090  
; REFERENCE/DOCKET NUMBER: Centro-2R  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908)530-6671  
; TELEFAX: (908)530-6584  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 112 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: Single  
; TOPOLOGY: linear  
; MOLECULE TYPE: "lysoil binding site" from E. coli  
; MOLECULE TYPE: Acetyl-transferase  
US-07-754-918A-12

Query Match 0.9%; Score 7; DB 1; Length 112;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 AAPAKQ 225  
Db 81 AAPAKQ 87

RESULT 26  
US-09-252-991A-20183  
; Sequence 20183, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A



```
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20183
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20183

Query Match          0.9%; Score 7; DB 4; Length 124;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 SAPARPA 138
Db 55 SAPARPA 61

RESULT 27
US-09-485-316A-7
; Sequence 7, Application US/09485316A
; Patent No. 634441
; GENERAL INFORMATION:
; APPLICANT: Bihaun, Bernard
; APPLICANT: Bouqueleret, Lydie
; APPLICANT: Yen-Potin, Frances
; TITLE OF INVENTION: Lipoprotein-regulating medicaments
; FILE REFERENCE: GENSET.036APC
; CURRENT APPLICATION NUMBER: US/09/485.316A
; CURRENT FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: FR 97/10088
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: FR 98/05032
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: PCT IB98/01256
; PRIOR FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent.pm
; SEQ ID NO 7
; LENGTH: 129
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: 1..129
; OTHER INFORMATION: fragment 117..245 of ref swissprot P02745
US-09-485-316A-7

Query Match          0.9%; Score 7; DB 4; Length 129;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 319 YFTFQVL 325
Db 44 YFTFQVL 50

RESULT 28
US-09-252-991A-28069
; Sequence 28069, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25515
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25515

Query Match          0.9%; Score 7; DB 4; Length 154;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPA 147
Db 121 AAPAAPA 127

RESULT 30
US-09-252-991A-25515
; Sequence 25515, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25515
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25515

Query Match          0.9%; Score 7; DB 4; Length 154;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 QQAAPA 223  
 Db 121 QQAAPA 127

## RESULT 31

US-08-775-978-1  
 ; Sequence 1, Application US/08775978  
 ; Patent No. 5831049  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HILLMAN, JENNIFER L.  
 ; TITLE OF INVENTION: NOVEL HUMAN THIOREDOXIN  
 ; NUMBER OF SEQUENCES: 3  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/775,978  
 ; FILING DATE: To Be Assigned  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Billings, Lucy J.  
 ; REGISTRATION NUMBER: 36,749  
 ; REFERENCE/DOCKET NUMBER: PF-0176 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-855-0555  
 ; TELEFAX: 415-845-4166  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 166 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; CLONE: 1430906  
 ; US-08-775-978-1

Query Match 0.9%; Score 7; DB 2; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 655 GLTVTPN 661  
 Db 41 GLTVTPN 47

## RESULT 32

US-09-252-991A-23675  
 ; Sequence 23675, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 23675  
 ; LENGTH: 179  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-23675

Query Match 0.9%; Score 7; DB 4; Length 179;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPA 147  
 Db 63 AAPAAPA 69

## RESULT 33

US-09-489-847-314  
 ; Sequence 314, Application US/09489847  
 ; Patent No. 6476195  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al  
 ; TITLE OF INVENTION: 98 Human Secreted Proteins  
 ; FILE REFERENCE: PZ031P1  
 ; CURRENT APPLICATION NUMBER: US/09/489,847  
 ; CURRENT FILING DATE: 2000-01-24  
 ; EARLIER APPLICATION NUMBER: PCT/US99/17130  
 ; EARLIER FILING DATE: 1999-07-29  
 ; EARLIER APPLICATION NUMBER: 60/094,657  
 ; EARLIER FILING DATE: 1998-07-30  
 ; EARLIER APPLICATION NUMBER: 60/095,486  
 ; EARLIER FILING DATE: 1998-08-05  
 ; EARLIER APPLICATION NUMBER: 60/096,319  
 ; EARLIER FILING DATE: 1998-08-12  
 ; EARLIER APPLICATION NUMBER: 60/095,454  
 ; EARLIER FILING DATE: 1998-08-06  
 ; EARLIER APPLICATION NUMBER: 60/095,455  
 ; EARLIER FILING DATE: 1998-08-06  
 ; NUMBER OF SEQ ID NOS: 376  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 314  
 ; LENGTH: 188  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-489-847-314

Query Match 0.9%; Score 7; DB 4; Length 188;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 QTAAPAK 208  
 Db 80 QTAAPAK 86

## RESULT 34

US-09-252-991A-30359  
 ; Sequence 30359, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142

```

; SEQ ID NO 30359
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30359

Query Match      0.9%; Score 7; DB 4; Length 197;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 SAPARPA 138
Db 180 SAPARPA 186

RESULT 35
US-09-107-532A-4916
; Sequence 4916, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/885,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4916:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 216 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...216
; SEQUENCE DESCRIPTION: SEQ ID NO: 4916:
US-09-107-532A-4916

Query Match      0.9%; Score 7; DB 4; Length 216;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 141 AAPARPA 147
Db 200 AAPARPA 206

; SEQ ID NO 30359
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30359

Query Match      0.9%; Score 7; DB 4; Length 197;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 SAPARPA 138
Db 180 SAPARPA 186

RESULT 36
US-09-041-889-4
; Sequence 4, Application US/09041899
; Patent No. 6033864
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Cohavy, Offer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
; TITLE OF INVENTION: Microbial UC PANCA antigens
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,889
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/837,058
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1...218
; OTHER INFORMATION: /note= "product = Human Histone
; OTHER INFORMATION: H1-S-4"
US-09-041-889-4

Query Match      0.9%; Score 7; DB 3; Length 218;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 141 AAPARPA 147
Db 6 AAPARPA 12

RESULT 37
US-08-837-058-4
; Sequence 4, Application US/08837058
; Patent No. 6074835
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Targan, Stephan R.
; APPLICANT: Eggena, Mark
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
; TITLE OF INVENTION: Histone H1
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/837,058
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 2438
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..218
; OTHER INFORMATION: /note= "product = Human Histone
; US-08-837-058-4

Query Match 0.9%; Score 7; DB 3; Length 218;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 141 AAPAAPA 147
Db 6 AAPAAPA 12

RESULT 38
US-09-417-264-4
; Sequence 4, Application US/09417264
; Patent No. 6537768
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Cohavy, Offer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
; TITLE OF INVENTION: Microbial UC panCA antigens
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/417,264
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/041,889
; FILING DATE:

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..218
; OTHER INFORMATION: /note= "product = Human Histone
; US-09-417-264-4

Query Match 0.9%; Score 7; DB 4; Length 218;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 141 AAPAAPA 147
Db 6 AAPAAPA 12

RESULT 39
US-09-252-991A-18193
; Sequence 18193, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18193
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-18193

Query Match 0.9%; Score 7; DB 4; Length 221;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 133 APAREAV 139
Db 3 APAREAV 9

RESULT 40
US-09-252-991A-27628
; Sequence 27628, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190

```

```
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27628
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27628

Query Match      0.9%  Score 7; DB 4; Length 222;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      58 TSSPARI 64
      |||||
Db      123 TSSPARI 129

Search completed: December 9, 2003, 10:36:33
Job time : 28 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 10:36:19 ; Search time 37 Seconds  
(without alignments)  
3865.444 Million cell updates/sec

Title: US-09-701-271a-2

Perfect score: 769

Sequence: 1 MNYKLTIIISGLFVATAAFO.....ELLIFITPRIMTAGNSLRY 769

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 684280 seqs, 185983659 residues

Word size : 0

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep:\*

2: /cgn2\_6/ptodata/1/pubaa/PC1\_NEW\_PUB.pep:\*

3: /cgn2\_6/ptodata/1/pubaa/US06\_NEW\_PUB.pep:\*

4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep:\*

5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pep:\*

6: /cgn2\_6/ptodata/1/pubaa/PCUS\_PUBCOMB.pep:\*

7: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pep:\*

8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep:\*

9: /cgn2\_6/ptodata/1/pubaa/US09\_PUBCOMB.pep:\*

10: /cgn2\_6/ptodata/1/pubaa/US09\_PUBCOMB.pep:\*

11: /cgn2\_6/ptodata/1/pubaa/US09C\_PUBCOMB.pep:\*

12: /cgn2\_6/ptodata/1/pubaa/US09C\_NEW\_PUB.pep:\*

13: /cgn2\_6/ptodata/1/pubaa/US10A\_PUBCOMB.pep:\*

14: /cgn2\_6/ptodata/1/pubaa/US10B\_PUBCOMB.pep:\*

15: /cgn2\_6/ptodata/1/pubaa/US10C\_PUBCOMB.pep:\*

16: /cgn2\_6/ptodata/1/pubaa/US10\_NEW\_PUB.pep:\*

17: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pep:\*

18: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description         |
|------------|-------|-------------|--------|----|---------------------|
| 1          | 356   | 46.3        | 769    | 12 | US-10-320-800-10    |
| 2          | 9     | 1.2         | 654    | 9  | US-09-815-242-10361 |
| 3          | 8     | 1.0         | 44     | 15 | US-10-156-761-11589 |
| 4          | 8     | 1.0         | 776    | 9  | US-09-815-242-5092  |
| 5          | 8     | 1.0         | 1024   | 15 | US-10-211-962-49    |
| 6          | 8     | 1.0         | 2383   | 9  | US-09-912-020-302   |
| 7          | 7     | 0.9         | 7      | 15 | US-10-006-869-2439  |
| 8          | 7     | 0.9         | 10     | 11 | US-09-572-404B-2232 |
| 9          | 7     | 0.9         | 10     | 11 | US-09-572-404B-2234 |
| 10         | 7     | 0.9         | 10     | 11 | US-09-572-404B-2416 |
| 11         | 7     | 0.9         | 10     | 11 | US-09-572-404B-2418 |
| 12         | 7     | 0.9         | 19     | 15 | US-10-225-567A-1405 |
| 13         | 7     | 0.9         | 33     | 12 | US-10-300-083-36    |
| 14         | 7     | 0.9         | 51     | 12 | US-09-851-873-47    |
| 15         | 7     | 0.9         | 56     | 11 | US-09-764-891-5018  |

|    |   |     |     |    |                     |                    |
|----|---|-----|-----|----|---------------------|--------------------|
| 16 | 7 | 0.9 | 56  | 15 | US-10-205-428-430   | Sequence 430, Appl |
| 17 | 7 | 0.9 | 67  | 12 | US-10-029-386-30401 | Sequence 30401, A  |
| 18 | 7 | 0.9 | 71  | 10 | US-09-943-123-1     | Sequence 1, Appl   |
| 19 | 7 | 0.9 | 71  | 10 | US-09-943-123-2     | Sequence 2, Appl   |
| 20 | 7 | 0.9 | 78  | 11 | US-09-820-843A-89   | Sequence 89, Appl  |
| 21 | 7 | 0.9 | 94  | 10 | US-09-738-626-6042  | Sequence 6042, Ap  |
| 22 | 7 | 0.9 | 129 | 14 | US-10-072-159-7     | Sequence 7, Appl   |
| 23 | 7 | 0.9 | 129 | 14 | US-10-060-845-7     | Sequence 7, Appl   |
| 24 | 7 | 0.9 | 130 | 10 | US-09-893-737-144   | Sequence 144, App  |
| 25 | 7 | 0.9 | 157 | 15 | US-10-156-761-13774 | Sequence 13774, A  |
| 26 | 7 | 0.9 | 158 | 15 | US-10-156-761-10017 | Sequence 10017, A  |
| 27 | 7 | 0.9 | 166 | 12 | US-10-032-2018-153  | Sequence 153, App  |
| 28 | 7 | 0.9 | 167 | 15 | US-10-102-806-656   | Sequence 656, App  |
| 29 | 7 | 0.9 | 182 | 15 | US-10-050-704-331   | Sequence 331, App  |
| 30 | 7 | 0.9 | 182 | 15 | US-10-156-761-9729  | Sequence 9729, Ap  |
| 31 | 7 | 0.9 | 197 | 9  | US-09-815-242-5457  | Sequence 5457, Ap  |
| 32 | 7 | 0.9 | 198 | 9  | US-09-815-242-12680 | Sequence 12680, A  |
| 33 | 7 | 0.9 | 205 | 9  | US-09-764-898-187   | Sequence 187, App  |
| 34 | 7 | 0.9 | 205 | 10 | US-09-738-626-3591  | Sequence 3591, Ap  |
| 35 | 7 | 0.9 | 208 | 15 | US-10-156-761-13477 | Sequence 13477, A  |
| 36 | 7 | 0.9 | 217 | 9  | US-09-811-284-177   | Sequence 177, App  |
| 37 | 7 | 0.9 | 218 | 15 | US-10-229-567-4     | Sequence 4, Appl   |
| 38 | 7 | 0.9 | 223 | 15 | US-10-128-714-3122  | Sequence 3122, Ap  |
| 39 | 7 | 0.9 | 223 | 9  | US-09-970-711-11    | Sequence 11, Appl  |
| 40 | 7 | 0.9 | 229 | 9  | US-09-800-729-117   | Sequence 117, App  |
| 41 | 7 | 0.9 | 234 | 15 | US-10-262-209-1     | Sequence 1, Appl   |
| 42 | 7 | 0.9 | 245 | 9  | US-09-789-561-154   | Sequence 85, Appl  |
| 43 | 7 | 0.9 | 245 | 9  | US-09-800-729-85    | Sequence 115, App  |
| 44 | 7 | 0.9 | 245 | 9  | US-09-800-729-115   | Sequence 116, App  |
| 45 | 7 | 0.9 | 245 | 9  | US-09-800-729-118   | Sequence 118, App  |
| 46 | 7 | 0.9 | 245 | 9  | US-09-800-729-119   | Sequence 119, App  |
| 47 | 7 | 0.9 | 245 | 10 | US-09-911-1788-49   | Sequence 49, Appl  |
| 48 | 7 | 0.9 | 245 | 12 | US-10-360-186-52    | Sequence 52, Appl  |
| 49 | 7 | 0.9 | 245 | 12 | US-10-197-293-4     | Sequence 4, Appl   |
| 50 | 7 | 0.9 | 245 | 15 | US-10-180-763-52    | Sequence 60, Appl  |
| 51 | 7 | 0.9 | 245 | 15 | US-10-295-403-60    | Sequence 60, Appl  |
| 52 | 7 | 0.9 | 270 | 10 | US-09-738-626-6861  | Sequence 6861, Ap  |
| 53 | 7 | 0.9 | 284 | 15 | US-10-156-761-8340  | Sequence 8340, Ap  |
| 54 | 7 | 0.9 | 287 | 15 | US-10-128-714-8122  | Sequence 8122, Ap  |
| 55 | 7 | 0.9 | 296 | 15 | US-10-098-807-7     | Sequence 7, Appl   |
| 56 | 7 | 0.9 | 297 | 15 | US-10-156-761-11575 | Sequence 11575, A  |
| 57 | 7 | 0.9 | 301 | 15 | US-10-050-704-340   | Sequence 340, App  |
| 58 | 7 | 0.9 | 308 | 12 | US-10-306-762-139   | Sequence 139, App  |
| 59 | 7 | 0.9 | 319 | 15 | US-10-122-466A-14   | Sequence 14, Appl  |
| 60 | 7 | 0.9 | 322 | 9  | US-09-789-836-6     | Sequence 6, Appl   |
| 61 | 7 | 0.9 | 322 | 12 | US-09-789-831-6     | Sequence 6, Appl   |
| 62 | 7 | 0.9 | 340 | 11 | US-09-880-505-37    | Sequence 37, Appl  |
| 63 | 7 | 0.9 | 340 | 14 | US-10-051-643-37    | Sequence 37, Appl  |
| 64 | 7 | 0.9 | 345 | 9  | US-09-789-836-5     | Sequence 5, Appl   |
| 65 | 7 | 0.9 | 345 | 12 | US-09-789-831-5     | Sequence 5, Appl   |
| 66 | 7 | 0.9 | 345 | 12 | US-10-251-661-10    | Sequence 10, Appl  |
| 67 | 7 | 0.9 | 345 | 15 | US-10-060-036-183   | Sequence 183, Appl |
| 68 | 7 | 0.9 | 348 | 11 | US-09-880-505-170   | Sequence 170, Appl |
| 69 | 7 | 0.9 | 348 | 12 | US-10-205-979-39    | Sequence 39, Appl  |
| 70 | 7 | 0.9 | 348 | 14 | US-10-051-643-170   | Sequence 170, App  |
| 71 | 7 | 0.9 | 351 | 15 | US-10-156-761-11443 | Sequence 11443, A  |
| 72 | 7 | 0.9 | 351 | 12 | US-10-029-386-33377 | Sequence 33377, A  |
| 73 | 7 | 0.9 | 358 | 12 | US-09-882-227-164   | Sequence 164, App  |
| 74 | 7 | 0.9 | 362 | 15 | US-10-050-882-68    | Sequence 68, Appl  |
| 75 | 7 | 0.9 | 364 | 9  | US-09-051-755-14    | Sequence 14, Appl  |
| 76 | 7 | 0.9 | 364 | 15 | US-10-231-055-14    | Sequence 14, Appl  |
| 77 | 7 | 0.9 | 365 | 15 | US-10-187-536-4     | Sequence 4, Appl   |
| 78 | 7 | 0.9 | 370 | 10 | US-09-925-300-1483  | Sequence 1483, Ap  |
| 79 | 7 | 0.9 | 384 | 10 | US-09-814-777A-18   | Sequence 18, Appl  |
| 80 | 7 | 0.9 | 384 | 10 | US-09-814-777A-20   | Sequence 20, Appl  |
| 81 | 7 | 0.9 | 386 | 15 | US-10-147-951B-9    | Sequence 9, Appl   |
| 82 | 7 | 0.9 | 388 | 11 | US-09-952-680A-17   | Sequence 17, Appl  |
| 83 | 7 | 0.9 | 391 | 11 | US-09-990-940-20    | Sequence 20, Appl  |
| 84 | 7 | 0.9 | 391 | 11 | US-09-841-720-8     | Sequence 8, Appl   |
| 85 | 7 | 0.9 | 391 | 11 | US-09-964-923A-12   | Sequence 12, Appl  |
| 86 | 7 | 0.9 | 391 | 11 | US-09-964-923A-13   | Sequence 13, Appl  |
| 87 | 7 | 0.9 | 391 | 11 | US-09-964-923A-14   | Sequence 14, Appl  |
| 88 | 7 | 0.9 | 391 | 11 | US-09-964-923A-14   | Sequence 14, Appl  |

|     |   |     |      |    |                     |                   |     |   |     |      |    |                     |                   |
|-----|---|-----|------|----|---------------------|-------------------|-----|---|-----|------|----|---------------------|-------------------|
| 89  | 7 | 0.9 | 391  | 12 | US-09-826-509-569   | Sequence 569, App | 162 | 7 | 0.9 | 4685 | 15 | US-10-156-761-10433 | Sequence 10433, A |
| 90  | 7 | 0.9 | 391  | 12 | US-10-353-684A-20   | Sequence 20, Appl | 163 | 7 | 0.9 | 6291 | 12 | US-10-329-079-41    | Sequence 41, Appl |
| 91  | 7 | 0.9 | 391  | 15 | US-10-225-567A-312  | Sequence 312, App | 164 | 7 | 0.9 | 8360 | 12 | US-10-132-134-34    | Sequence 34, Appl |
| 92  | 7 | 0.9 | 393  | 11 | US-09-880-505-34    | Sequence 94, Appl | 165 | 6 | 0.8 | 7    | 15 | US-10-006-869-2549  | Sequence 2549, Ap |
| 93  | 7 | 0.9 | 393  | 14 | US-10-051-643-94    | Sequence 94, Appl | 166 | 6 | 0.8 | 8    | 15 | US-10-006-869-2430  | Sequence 2430, Ap |
| 94  | 7 | 0.9 | 394  | 10 | US-09-738-626-4096  | Sequence 4096, Ap | 167 | 6 | 0.8 | 8    | 15 | US-10-006-869-2432  | Sequence 2432, Ap |
| 95  | 7 | 0.9 | 394  | 15 | US-10-050-882-137   | Sequence 137, App | 168 | 6 | 0.8 | 8    | 15 | US-10-006-869-2550  | Sequence 2550, Ap |
| 96  | 7 | 0.9 | 403  | 10 | US-09-738-626-6488  | Sequence 6488, A  | 169 | 6 | 0.8 | 9    | 12 | US-10-024-652-1     | Sequence 1, Appl  |
| 97  | 7 | 0.9 | 408  | 15 | US-10-156-761-11222 | Sequence 11222, A | 170 | 6 | 0.8 | 9    | 12 | US-10-024-652-408   | Sequence 408, App |
| 98  | 7 | 0.9 | 419  | 15 | US-10-156-761-13549 | Sequence 13549, A | 171 | 6 | 0.8 | 9    | 12 | US-10-024-652-701   | Sequence 701, App |
| 99  | 7 | 0.9 | 432  | 10 | US-09-881-752A-198  | Sequence 198, App | 172 | 6 | 0.8 | 9    | 12 | US-10-024-652-705   | Sequence 705, App |
| 100 | 7 | 0.9 | 433  | 14 | US-10-078-929-172   | Sequence 172, App | 173 | 6 | 0.8 | 9    | 12 | US-10-024-652-707   | Sequence 707, App |
| 101 | 7 | 0.9 | 446  | 14 | US-10-078-929-52    | Sequence 52, Appl | 174 | 6 | 0.8 | 9    | 12 | US-10-024-652-721   | Sequence 721, App |
| 102 | 7 | 0.9 | 448  | 12 | US-10-017-161-1998  | Sequence 1998, Ap | 175 | 6 | 0.8 | 9    | 12 | US-10-024-652-723   | Sequence 723, App |
| 103 | 7 | 0.9 | 478  | 9  | US-09-813-555-1     | Sequence 1, Appl  | 176 | 6 | 0.8 | 9    | 12 | US-10-024-652-724   | Sequence 724, App |
| 104 | 7 | 0.9 | 478  | 9  | US-09-813-555-3     | Sequence 3, Appl  | 177 | 6 | 0.8 | 9    | 12 | US-10-024-652-740   | Sequence 740, App |
| 105 | 7 | 0.9 | 478  | 10 | US-09-943-123-6     | Sequence 6, Appl  | 178 | 6 | 0.8 | 9    | 12 | US-10-024-652-741   | Sequence 741, App |
| 106 | 7 | 0.9 | 478  | 10 | US-09-943-123-7     | Sequence 7, Appl  | 179 | 6 | 0.8 | 9    | 12 | US-10-024-652-745   | Sequence 745, App |
| 107 | 7 | 0.9 | 478  | 10 | US-09-943-123-12    | Sequence 11, Appl | 180 | 6 | 0.8 | 9    | 12 | US-10-024-652-759   | Sequence 759, App |
| 108 | 7 | 0.9 | 478  | 10 | US-09-943-123-13    | Sequence 13, Appl | 181 | 6 | 0.8 | 9    | 12 | US-10-024-652-761   | Sequence 761, App |
| 109 | 7 | 0.9 | 484  | 12 | US-10-402-842-43    | Sequence 43, Appl | 182 | 6 | 0.8 | 9    | 12 | US-10-024-652-765   | Sequence 765, App |
| 110 | 7 | 0.9 | 487  | 15 | US-10-050-704-335   | Sequence 335, App | 183 | 6 | 0.8 | 9    | 12 | US-10-024-652-777   | Sequence 777, App |
| 111 | 7 | 0.9 | 500  | 15 | US-10-106-698-4765  | Sequence 4765, Ap | 184 | 6 | 0.8 | 9    | 12 | US-10-024-652-781   | Sequence 781, App |
| 112 | 7 | 0.9 | 516  | 11 | US-09-291-417-6     | Sequence 6, Appl  | 185 | 6 | 0.8 | 9    | 12 | US-10-024-652-782   | Sequence 782, App |
| 113 | 7 | 0.9 | 526  | 15 | US-10-050-704-336   | Sequence 336, App | 186 | 6 | 0.8 | 9    | 12 | US-10-024-652-787   | Sequence 787, App |
| 114 | 7 | 0.9 | 539  | 9  | US-09-815-242-10497 | Sequence 10497, A | 187 | 6 | 0.8 | 9    | 12 | US-10-024-652-800   | Sequence 800, App |
| 115 | 7 | 0.9 | 547  | 10 | US-09-771-161A-272  | Sequence 272, App | 188 | 6 | 0.8 | 9    | 12 | US-10-024-652-801   | Sequence 801, App |
| 116 | 7 | 0.9 | 547  | 10 | US-09-771-161A-273  | Sequence 273, App | 189 | 6 | 0.8 | 9    | 12 | US-10-024-652-816   | Sequence 816, App |
| 117 | 7 | 0.9 | 547  | 11 | US-09-943-857-4     | Sequence 4, Appl  | 190 | 6 | 0.8 | 9    | 12 | US-10-024-652-818   | Sequence 818, App |
| 118 | 7 | 0.9 | 547  | 11 | US-09-943-857-8     | Sequence 8, Appl  | 191 | 6 | 0.8 | 9    | 12 | US-10-024-652-819   | Sequence 819, App |
| 119 | 7 | 0.9 | 547  | 11 | US-09-943-857-10    | Sequence 10, Appl | 192 | 6 | 0.8 | 9    | 12 | US-10-024-652-947   | Sequence 947, App |
| 120 | 7 | 0.9 | 551  | 12 | US-09-976-782-89    | Sequence 89, Appl | 193 | 6 | 0.8 | 9    | 12 | US-10-024-652-1276  | Sequence 1276, Ap |
| 121 | 7 | 0.9 | 551  | 12 | US-09-976-782-90    | Sequence 90, Appl | 194 | 6 | 0.8 | 9    | 12 | US-10-024-652-2068  | Sequence 2068, Ap |
| 122 | 7 | 0.9 | 551  | 12 | US-10-015-115-127   | Sequence 127, App | 195 | 6 | 0.8 | 9    | 12 | US-10-024-652-2083  | Sequence 2083, Ap |
| 123 | 7 | 0.9 | 561  | 9  | US-09-815-242-11673 | Sequence 11673, A | 196 | 6 | 0.8 | 9    | 12 | US-10-024-652-2086  | Sequence 2086, Ap |
| 124 | 7 | 0.9 | 572  | 10 | US-09-986-632-4     | Sequence 4, Appl  | 197 | 6 | 0.8 | 9    | 15 | US-10-006-869-2431  | Sequence 2431, Ap |
| 125 | 7 | 0.9 | 578  | 12 | US-10-032-585-7793  | Sequence 7793, Ap | 198 | 6 | 0.8 | 9    | 15 | US-10-006-869-2435  | Sequence 2435, Ap |
| 126 | 7 | 0.9 | 578  | 12 | US-10-029-386-33840 | Sequence 33840, A | 199 | 6 | 0.8 | 9    | 15 | US-10-006-869-2551  | Sequence 2551, Ap |
| 127 | 7 | 0.9 | 603  | 15 | US-10-270-333-189   | Sequence 189, App | 200 | 6 | 0.8 | 9    | 15 | US-10-235-483-43    | Sequence 43, Appl |
| 128 | 7 | 0.9 | 610  | 14 | US-10-108-605-333   | Sequence 333, App | 201 | 6 | 0.8 | 10   | 11 | US-09-572-404B-2583 | Sequence 2583, Ap |
| 129 | 7 | 0.9 | 612  | 15 | US-10-156-761-10209 | Sequence 10209, A | 202 | 6 | 0.8 | 10   | 12 | US-09-572-270A-441  | Sequence 441, App |
| 130 | 7 | 0.9 | 614  | 15 | US-10-145-012-2     | Sequence 2, Appl  | 203 | 6 | 0.8 | 10   | 12 | US-09-572-270A-443  | Sequence 443, App |
| 131 | 7 | 0.9 | 614  | 15 | US-10-145-012-13    | Sequence 13, Appl | 204 | 6 | 0.8 | 10   | 12 | US-09-572-270A-445  | Sequence 445, App |
| 132 | 7 | 0.9 | 616  | 15 | US-10-156-761-13507 | Sequence 13507, A | 205 | 6 | 0.8 | 10   | 12 | US-10-024-652-70    | Sequence 70, Appl |
| 133 | 7 | 0.9 | 627  | 14 | US-10-047-260-36    | Sequence 36, Appl | 206 | 6 | 0.8 | 10   | 12 | US-10-024-652-97    | Sequence 97, Appl |
| 134 | 7 | 0.9 | 630  | 9  | US-09-815-242-10036 | Sequence 10036, A | 207 | 6 | 0.8 | 10   | 12 | US-10-024-652-184   | Sequence 184, App |
| 135 | 7 | 0.9 | 637  | 12 | US-10-294-835-2     | Sequence 2, Appl  | 208 | 6 | 0.8 | 10   | 12 | US-10-024-652-257   | Sequence 257, App |
| 136 | 7 | 0.9 | 643  | 10 | US-10-234-835-4     | Sequence 4, Appl  | 209 | 6 | 0.8 | 10   | 12 | US-10-024-652-278   | Sequence 278, App |
| 137 | 7 | 0.9 | 643  | 10 | US-09-943-671-33    | Sequence 33, Appl | 210 | 6 | 0.8 | 10   | 12 | US-10-024-652-294   | Sequence 294, App |
| 138 | 7 | 0.9 | 649  | 15 | US-10-078-770-192   | Sequence 192, App | 211 | 6 | 0.8 | 10   | 12 | US-10-024-652-694   | Sequence 694, App |
| 139 | 7 | 0.9 | 653  | 15 | US-10-156-761-10327 | Sequence 10327, A | 212 | 6 | 0.8 | 10   | 12 | US-10-024-652-710   | Sequence 710, App |
| 140 | 7 | 0.9 | 673  | 15 | US-10-156-761-8864  | Sequence 8864, Ap | 213 | 6 | 0.8 | 10   | 12 | US-10-024-652-711   | Sequence 711, App |
| 141 | 7 | 0.9 | 681  | 9  | US-09-881-736-6     | Sequence 6, Appl  | 214 | 6 | 0.8 | 10   | 12 | US-10-024-652-717   | Sequence 717, App |
| 142 | 7 | 0.9 | 749  | 10 | US-09-833-205-4     | Sequence 4, Appl  | 215 | 6 | 0.8 | 10   | 12 | US-10-024-652-718   | Sequence 718, App |
| 143 | 7 | 0.9 | 757  | 12 | US-10-168-445-15    | Sequence 15, Appl | 216 | 6 | 0.8 | 10   | 12 | US-10-024-652-729   | Sequence 729, App |
| 144 | 7 | 0.9 | 861  | 11 | US-09-820-843A-109  | Sequence 109, App | 217 | 6 | 0.8 | 10   | 12 | US-10-024-652-732   | Sequence 732, App |
| 145 | 7 | 0.9 | 904  | 9  | US-09-894-998-18    | Sequence 18, Appl | 218 | 6 | 0.8 | 10   | 12 | US-10-024-652-734   | Sequence 734, App |
| 146 | 7 | 0.9 | 904  | 12 | US-10-200-562-18    | Sequence 18, Appl | 219 | 6 | 0.8 | 10   | 12 | US-10-024-652-735   | Sequence 735, App |
| 147 | 7 | 0.9 | 904  | 15 | US-10-237-551-18    | Sequence 18, Appl | 220 | 6 | 0.8 | 10   | 12 | US-10-024-652-748   | Sequence 748, App |
| 148 | 7 | 0.9 | 904  | 15 | US-10-121-988-18    | Sequence 18, Appl | 221 | 6 | 0.8 | 10   | 12 | US-10-024-652-752   | Sequence 752, App |
| 149 | 7 | 0.9 | 930  | 15 | US-10-300-834-6     | Sequence 6, Appl  | 222 | 6 | 0.8 | 10   | 12 | US-10-024-652-753   | Sequence 753, App |
| 150 | 7 | 0.9 | 934  | 15 | US-10-128-714-3368  | Sequence 3368, Ap | 223 | 6 | 0.8 | 10   | 12 | US-10-024-652-755   | Sequence 755, App |
| 151 | 7 | 0.9 | 959  | 15 | US-10-128-714-8388  | Sequence 8388, Ap | 224 | 6 | 0.8 | 10   | 12 | US-10-024-652-767   | Sequence 767, App |
| 152 | 7 | 0.9 | 1005 | 15 | US-10-059-962-4     | Sequence 4, Appl  | 225 | 6 | 0.8 | 10   | 12 | US-10-024-652-774   | Sequence 774, App |
| 153 | 7 | 0.9 | 1095 | 12 | US-10-193-764-65    | Sequence 65, Appl | 226 | 6 | 0.8 | 10   | 12 | US-10-024-652-776   | Sequence 776, App |
| 154 | 7 | 0.9 | 1276 | 15 | US-10-156-761-10509 | Sequence 10509, A | 227 | 6 | 0.8 | 10   | 12 | US-10-024-652-786   | Sequence 786, App |
| 155 | 7 | 0.9 | 1373 | 12 | US-10-032-585-7129  | Sequence 7129, Ap | 228 | 6 | 0.8 | 10   | 12 | US-10-024-652-787   | Sequence 787, App |
| 156 | 7 | 0.9 | 1536 | 14 | US-10-193-764-63    | Sequence 63, Appl | 229 | 6 | 0.8 | 10   | 12 | US-10-024-652-790   | Sequence 790, App |
| 157 | 7 | 0.9 | 1536 | 14 | US-10-092-880-2     | Sequence 2, Appl  | 230 | 6 | 0.8 | 10   | 12 | US-10-024-652-794   | Sequence 794, App |
| 158 | 7 | 0.9 | 1573 | 12 | US-10-214-766-35    | Sequence 35, Appl | 231 | 6 | 0.8 | 10   | 12 | US-10-024-652-807   | Sequence 807, App |
| 159 | 7 | 0.9 | 1633 | 12 | US-10-029-386-33090 | Sequence 33090, A | 232 | 6 | 0.8 | 10   | 12 | US-10-024-652-809   | Sequence 809, App |
| 160 | 7 | 0.9 | 1929 | 14 | US-10-108-605-347   | Sequence 347, App | 233 | 6 | 0.8 | 10   | 12 | US-10-024-652-811   | Sequence 811, App |
| 161 | 7 | 0.9 | 2069 | 10 | US-09-738-626-4320  | Sequence 4320, Ap | 234 | 6 | 0.8 | 10   | 12 | US-10-024-652-813   | Sequence 813, App |

|     |   |     |    |    |                     |                    |     |   |     |    |    |                     |                    |
|-----|---|-----|----|----|---------------------|--------------------|-----|---|-----|----|----|---------------------|--------------------|
| 235 | 6 | 0.8 | 10 | 12 | US-10-024-652-824   | Sequence 824, App  | 308 | 6 | 0.8 | 23 | 15 | US-10-059-261-91    | Sequence 91, Appl  |
| 236 | 6 | 0.8 | 10 | 12 | US-10-024-652-827   | Sequence 827, App  | 309 | 6 | 0.8 | 23 | 15 | US-10-059-261-199   | Sequence 199, App  |
| 237 | 6 | 0.8 | 10 | 12 | US-10-024-652-831   | Sequence 831, App  | 310 | 6 | 0.8 | 23 | 15 | US-10-057-789-196   | Sequence 196, App  |
| 238 | 6 | 0.8 | 10 | 12 | US-10-024-652-832   | Sequence 832, App  | 311 | 6 | 0.8 | 23 | 15 | US-10-121-628-196   | Sequence 196, App  |
| 239 | 6 | 0.8 | 10 | 12 | US-10-024-652-1601  | Sequence 1601, App | 312 | 6 | 0.8 | 23 | 15 | US-10-120-604-221   | Sequence 221, App  |
| 240 | 6 | 0.8 | 10 | 12 | US-10-024-652-1661  | Sequence 1661, App | 313 | 6 | 0.8 | 23 | 15 | US-10-106-698-7753  | Sequence 7753, App |
| 241 | 6 | 0.8 | 10 | 12 | US-10-024-652-1712  | Sequence 1712, App | 314 | 6 | 0.8 | 23 | 9  | US-09-864-761-46567 | Sequence 46567, A  |
| 242 | 6 | 0.8 | 10 | 12 | US-10-024-652-1732  | Sequence 1732, App | 315 | 6 | 0.8 | 23 | 12 | US-09-988-1658-10   | Sequence 10, Appl  |
| 243 | 6 | 0.8 | 10 | 12 | US-10-024-652-1747  | Sequence 1747, App | 316 | 6 | 0.8 | 23 | 11 | US-09-809-391-660   | Sequence 660, App  |
| 244 | 6 | 0.8 | 10 | 12 | US-10-024-652-1820  | Sequence 1820, App | 317 | 6 | 0.8 | 23 | 12 | US-09-882-171-660   | Sequence 660, App  |
| 245 | 6 | 0.8 | 10 | 12 | US-10-024-652-1830  | Sequence 1830, App | 318 | 6 | 0.8 | 33 | 12 | US-10-029-386-32000 | Sequence 32000, A  |
| 246 | 6 | 0.8 | 10 | 12 | US-10-024-652-2088  | Sequence 2088, App | 319 | 6 | 0.8 | 34 | 9  | US-09-864-761-36214 | Sequence 36214, A  |
| 247 | 6 | 0.8 | 10 | 12 | US-10-024-652-2089  | Sequence 2089, App | 320 | 6 | 0.8 | 34 | 12 | US-10-289-135A-45   | Sequence 45, Appl  |
| 248 | 6 | 0.8 | 10 | 12 | US-10-024-652-2091  | Sequence 2091, App | 321 | 6 | 0.8 | 34 | 12 | US-10-395-897-8     | Sequence 8, Appl   |
| 249 | 6 | 0.8 | 10 | 12 | US-10-024-652-2092  | Sequence 2092, App | 322 | 6 | 0.8 | 34 | 12 | US-10-395-897-9     | Sequence 9, Appl   |
| 250 | 6 | 0.8 | 10 | 15 | US-10-235-483-42    | Sequence 42, Appl  | 323 | 6 | 0.8 | 35 | 10 | US-09-884-441-413   | Sequence 413, App  |
| 251 | 6 | 0.8 | 10 | 15 | US-10-146-938-2     | Sequence 2, Appl   | 324 | 6 | 0.8 | 35 | 11 | US-09-907-969-413   | Sequence 413, App  |
| 252 | 6 | 0.8 | 11 | 11 | US-09-775-938-41    | Sequence 41, Appl  | 325 | 6 | 0.8 | 35 | 12 | US-09-827-271-413   | Sequence 413, App  |
| 253 | 6 | 0.8 | 11 | 15 | US-10-146-938-4     | Sequence 4, Appl   | 326 | 6 | 0.8 | 35 | 15 | US-10-198-053-413   | Sequence 413, App  |
| 254 | 6 | 0.8 | 13 | 15 | US-10-235-483-40    | Sequence 40, Appl  | 327 | 6 | 0.8 | 35 | 15 | US-10-198-053-413   | Sequence 413, App  |
| 255 | 6 | 0.8 | 14 | 12 | US-10-005-876A-78   | Sequence 78, Appl  | 328 | 6 | 0.8 | 36 | 9  | US-09-864-761-48763 | Sequence 48763, A  |
| 256 | 6 | 0.8 | 14 | 15 | US-10-235-483-39    | Sequence 39, Appl  | 329 | 6 | 0.8 | 36 | 15 | US-10-026-741-85    | Sequence 85, Appl  |
| 257 | 6 | 0.8 | 15 | 12 | US-10-024-652-2116  | Sequence 2116, App | 330 | 6 | 0.8 | 40 | 12 | US-10-097-111-492   | Sequence 492, App  |
| 258 | 6 | 0.8 | 15 | 12 | US-10-024-652-2134  | Sequence 2134, App | 331 | 6 | 0.8 | 40 | 12 | US-10-315-964A-15   | Sequence 15, Appl  |
| 259 | 6 | 0.8 | 15 | 12 | US-10-024-652-2189  | Sequence 2189, App | 332 | 6 | 0.8 | 40 | 12 | US-10-317-251A-15   | Sequence 15, Appl  |
| 260 | 6 | 0.8 | 15 | 12 | US-10-024-652-2213  | Sequence 2213, App | 333 | 6 | 0.8 | 40 | 12 | US-10-317-252A-15   | Sequence 15, Appl  |
| 261 | 6 | 0.8 | 15 | 12 | US-10-024-652-2221  | Sequence 2221, App | 334 | 6 | 0.8 | 42 | 9  | US-09-864-761-40443 | Sequence 40443, A  |
| 262 | 6 | 0.8 | 15 | 12 | US-10-024-652-2233  | Sequence 2233, App | 335 | 6 | 0.8 | 42 | 9  | US-09-947-925A-12   | Sequence 12, Appl  |
| 263 | 6 | 0.8 | 15 | 12 | US-10-024-652-2285  | Sequence 2285, App | 336 | 6 | 0.8 | 44 | 9  | US-09-947-925A-13   | Sequence 13, Appl  |
| 264 | 6 | 0.8 | 15 | 12 | US-10-024-652-2351  | Sequence 2351, App | 337 | 6 | 0.8 | 44 | 9  | US-09-864-761-48991 | Sequence 48991, A  |
| 265 | 6 | 0.8 | 15 | 12 | US-10-024-652-2352  | Sequence 2352, App | 338 | 6 | 0.8 | 46 | 9  | US-09-864-761-36286 | Sequence 36286, A  |
| 266 | 6 | 0.8 | 15 | 12 | US-10-024-652-2393  | Sequence 2393, App | 339 | 6 | 0.8 | 51 | 9  | US-09-864-761-46744 | Sequence 46744, A  |
| 267 | 6 | 0.8 | 15 | 12 | US-10-024-652-2521  | Sequence 2521, App | 340 | 6 | 0.8 | 51 | 12 | US-09-864-761-47871 | Sequence 47871, A  |
| 268 | 6 | 0.8 | 15 | 12 | US-10-024-652-2553  | Sequence 2553, App | 341 | 6 | 0.8 | 52 | 9  | US-10-029-386-32874 | Sequence 32874, A  |
| 269 | 6 | 0.8 | 15 | 12 | US-10-024-652-2554  | Sequence 2554, App | 342 | 6 | 0.8 | 52 | 9  | US-09-864-761-35374 | Sequence 35374, A  |
| 270 | 6 | 0.8 | 15 | 12 | US-10-024-652-2555  | Sequence 2555, App | 343 | 6 | 0.8 | 53 | 9  | US-09-864-761-35370 | Sequence 35370, A  |
| 271 | 6 | 0.8 | 15 | 12 | US-10-024-652-2557  | Sequence 2557, App | 344 | 6 | 0.8 | 54 | 9  | US-09-925-297-466   | Sequence 466, App  |
| 272 | 6 | 0.8 | 15 | 12 | US-10-024-652-2558  | Sequence 2558, App | 345 | 6 | 0.8 | 57 | 9  | US-08-722-570-3     | Sequence 3, Appl   |
| 273 | 6 | 0.8 | 15 | 12 | US-10-024-652-2559  | Sequence 2559, App | 346 | 6 | 0.8 | 57 | 9  | US-09-864-761-46691 | Sequence 46691, A  |
| 274 | 6 | 0.8 | 15 | 12 | US-10-024-652-2561  | Sequence 2561, App | 347 | 6 | 0.8 | 57 | 12 | US-10-029-386-28701 | Sequence 28701, A  |
| 275 | 6 | 0.8 | 15 | 12 | US-10-024-652-2562  | Sequence 2562, App | 348 | 6 | 0.8 | 58 | 9  | US-09-764-869-886   | Sequence 886, App  |
| 276 | 6 | 0.8 | 15 | 12 | US-10-024-652-2563  | Sequence 2563, App | 349 | 6 | 0.8 | 58 | 15 | US-10-091-504-886   | Sequence 886, App  |
| 277 | 6 | 0.8 | 15 | 12 | US-10-024-652-2565  | Sequence 2565, App | 350 | 6 | 0.8 | 59 | 9  | US-09-864-761-41467 | Sequence 41467, A  |
| 278 | 6 | 0.8 | 15 | 12 | US-10-024-652-2567  | Sequence 2567, App | 351 | 6 | 0.8 | 59 | 12 | US-10-012-952A-159  | Sequence 159, App  |
| 279 | 6 | 0.8 | 15 | 15 | US-10-137-745-7     | Sequence 7, Appl   | 352 | 6 | 0.8 | 59 | 15 | US-10-100-679-77    | Sequence 77, Appl  |
| 280 | 6 | 0.8 | 15 | 15 | US-10-235-483-38    | Sequence 38, Appl  | 353 | 6 | 0.8 | 60 | 9  | US-09-864-761-40431 | Sequence 40431, A  |
| 281 | 6 | 0.8 | 15 | 15 | US-10-235-567A-1336 | Sequence 1336, App | 354 | 6 | 0.8 | 61 | 9  | US-09-864-761-48036 | Sequence 48036, A  |
| 282 | 6 | 0.8 | 16 | 9  | US-09-799-576A-19   | Sequence 19, Appl  | 355 | 6 | 0.8 | 61 | 15 | US-10-156-761-13958 | Sequence 13958, A  |
| 283 | 6 | 0.8 | 16 | 9  | US-09-799-576A-20   | Sequence 20, Appl  | 356 | 6 | 0.8 | 62 | 10 | US-09-764-847-572   | Sequence 572, App  |
| 284 | 6 | 0.8 | 16 | 9  | US-09-799-540-19    | Sequence 19, Appl  | 357 | 6 | 0.8 | 62 | 15 | US-10-093-154-572   | Sequence 572, App  |
| 285 | 6 | 0.8 | 16 | 9  | US-09-799-540-20    | Sequence 20, Appl  | 358 | 6 | 0.8 | 64 | 9  | US-09-864-761-48364 | Sequence 48364, A  |
| 286 | 6 | 0.8 | 16 | 10 | US-09-799-608-19    | Sequence 19, Appl  | 359 | 6 | 0.8 | 64 | 12 | US-10-029-386-32301 | Sequence 32301, A  |
| 287 | 6 | 0.8 | 16 | 10 | US-09-799-608-20    | Sequence 20, Appl  | 360 | 6 | 0.8 | 68 | 11 | US-09-764-891-3350  | Sequence 3350, App |
| 288 | 6 | 0.8 | 16 | 10 | US-09-822-873-4     | Sequence 4, Appl   | 361 | 6 | 0.8 | 68 | 15 | US-10-091-572-239   | Sequence 239, App  |
| 289 | 6 | 0.8 | 16 | 12 | US-10-840-421-3     | Sequence 3, Appl   | 362 | 6 | 0.8 | 69 | 10 | US-09-916-510A-6    | Sequence 6, Appl   |
| 290 | 6 | 0.8 | 16 | 12 | US-09-798-932-19    | Sequence 6, Appl   | 363 | 6 | 0.8 | 69 | 12 | US-10-029-386-31265 | Sequence 31265, A  |
| 291 | 6 | 0.8 | 18 | 12 | US-09-798-932-20    | Sequence 20, Appl  | 364 | 6 | 0.8 | 70 | 10 | US-09-848-664-9     | Sequence 9, Appl   |
| 292 | 6 | 0.8 | 18 | 12 | US-10-348-190-6     | Sequence 6, Appl   | 365 | 6 | 0.8 | 71 | 9  | US-09-864-761-43475 | Sequence 43475, A  |
| 293 | 6 | 0.8 | 20 | 10 | US-09-884-441-396   | Sequence 396, App  | 366 | 6 | 0.8 | 74 | 12 | US-10-084-843-291   | Sequence 291, App  |
| 294 | 6 | 0.8 | 20 | 11 | US-09-907-969-396   | Sequence 396, App  | 367 | 6 | 0.8 | 74 | 12 | US-10-193-002-286   | Sequence 286, App  |
| 295 | 6 | 0.8 | 20 | 12 | US-09-827-271-396   | Sequence 396, App  | 368 | 6 | 0.8 | 75 | 9  | US-09-925-301-849   | Sequence 849, App  |
| 296 | 6 | 0.8 | 20 | 12 | US-09-962-756-1771  | Sequence 1771, App | 369 | 6 | 0.8 | 76 | 11 | US-09-764-877-1597  | Sequence 1597, App |
| 297 | 6 | 0.8 | 20 | 12 | US-09-862-756-1909  | Sequence 1909, App | 370 | 6 | 0.8 | 76 | 11 | US-09-877-843-88    | Sequence 88, Appl  |
| 298 | 6 | 0.8 | 20 | 15 | US-10-198-053-396   | Sequence 396, App  | 371 | 6 | 0.8 | 78 | 11 | US-09-764-872-414   | Sequence 414, App  |
| 299 | 6 | 0.8 | 21 | 15 | US-10-198-053-602   | Sequence 602, App  | 372 | 6 | 0.8 | 79 | 9  | US-09-864-761-34986 | Sequence 34986, A  |
| 300 | 6 | 0.8 | 21 | 11 | US-09-880-505-13    | Sequence 13, Appl  | 373 | 6 | 0.8 | 79 | 15 | US-10-029-386-31939 | Sequence 31939, A  |
| 301 | 6 | 0.8 | 21 | 12 | US-09-930-915A-20   | Sequence 20, Appl  | 374 | 6 | 0.8 | 80 | 9  | US-09-764-870-510   | Sequence 510, App  |
| 302 | 6 | 0.8 | 21 | 12 | US-10-082-014-42    | Sequence 42, Appl  | 375 | 6 | 0.8 | 80 | 9  | US-09-864-761-46562 | Sequence 46562, A  |
| 303 | 6 | 0.8 | 21 | 12 | US-10-372-076-43    | Sequence 43, Appl  | 376 | 6 | 0.8 | 80 | 9  | US-09-864-761-47270 | Sequence 47270, A  |
| 304 | 6 | 0.8 | 21 | 14 | US-10-124-800-15    | Sequence 15, Appl  | 377 | 6 | 0.8 | 80 | 10 | US-09-764-847-626   | Sequence 626, App  |
| 305 | 6 | 0.8 | 21 | 14 | US-10-051-643-13    | Sequence 13, Appl  | 378 | 6 | 0.8 | 80 | 15 | US-10-092-154-626   | Sequence 626, App  |
| 306 | 6 | 0.8 | 22 | 12 | US-10-029-386-32170 | Sequence 32170, A  | 379 | 6 | 0.8 | 80 | 15 | US-10-135-540-510   | Sequence 510, App  |
| 307 | 6 | 0.8 | 23 | 15 | US-10-115-615-36    | Sequence 36, Appl  | 380 | 6 | 0.8 | 81 | 15 | US-10-156-761-10656 | Sequence 10656, A  |



|     |   |     |    |    |                     |                    |     |    |    |                   |                    |
|-----|---|-----|----|----|---------------------|--------------------|-----|----|----|-------------------|--------------------|
| 381 | 6 | 0.8 | 83 | 11 | US-09-764-891-3794  | Sequence 3794, Ap  | 454 | 95 | 12 | US-10-152-531-62  | Sequence 62, Appl  |
| 382 | 6 | 0.8 | 84 | 10 | US-09-925-300-1326  | Sequence 1326, Ap  | 455 | 95 | 12 | US-09-876-997-329 | Sequence 329, Appl |
| 383 | 6 | 0.8 | 85 | 10 | US-09-925-300-1078  | Sequence 1078, Ap  | 456 | 95 | 12 | US-10-127-840A-62 | Sequence 62, Appl  |
| 384 | 6 | 0.8 | 89 | 12 | US-10-115-223-28    | Sequence 28, Appl  | 457 | 95 | 12 | US-10-142-424-62  | Sequence 62, Appl  |
| 385 | 6 | 0.8 | 89 | 12 | US-10-166-087-22    | Sequence 22, Appl  | 458 | 95 | 12 | US-10-213-022-8   | Sequence 8, Appl   |
| 386 | 6 | 0.8 | 90 | 12 | US-10-074-566-119   | Sequence 119, Appl | 459 | 95 | 12 | US-10-142-763-62  | Sequence 62, Appl  |
| 387 | 6 | 0.8 | 90 | 12 | US-10-074-566-120   | Sequence 120, Appl | 460 | 95 | 12 | US-10-142-765-62  | Sequence 62, Appl  |
| 388 | 6 | 0.8 | 91 | 9  | US-09-864-761-43033 | Sequence 43033, A  | 461 | 95 | 12 | US-10-142-887-62  | Sequence 62, Appl  |
| 389 | 6 | 0.8 | 91 | 9  | US-09-864-761-43734 | Sequence 43734, A  | 462 | 95 | 12 | US-10-142-888-62  | Sequence 62, Appl  |
| 390 | 6 | 0.8 | 91 | 11 | US-09-764-891-2765  | Sequence 2765, Ap  | 463 | 95 | 12 | US-10-143-034-62  | Sequence 62, Appl  |
| 391 | 6 | 0.8 | 91 | 11 | US-09-764-891-4257  | Sequence 4257, Ap  | 464 | 95 | 12 | US-10-143-116-62  | Sequence 62, Appl  |
| 392 | 6 | 0.8 | 93 | 9  | US-09-864-761-37510 | Sequence 37510, A  | 465 | 95 | 12 | US-10-144-957-62  | Sequence 62, Appl  |
| 393 | 6 | 0.8 | 94 | 9  | US-09-734-569-120   | Sequence 120, Appl | 466 | 95 | 12 | US-10-144-952-62  | Sequence 62, Appl  |
| 394 | 6 | 0.8 | 94 | 10 | US-09-796-692-2363  | Sequence 2363, Ap  | 467 | 95 | 12 | US-10-145-045-62  | Sequence 62, Appl  |
| 395 | 6 | 0.8 | 94 | 15 | US-10-040-862-2363  | Sequence 2363, Ap  | 468 | 95 | 12 | US-10-145-080-62  | Sequence 62, Appl  |
| 396 | 6 | 0.8 | 95 | 10 | US-09-731-873-339   | Sequence 339, Appl | 469 | 95 | 12 | US-10-145-091-62  | Sequence 62, Appl  |
| 397 | 6 | 0.8 | 95 | 12 | US-10-137-870-62    | Sequence 62, Appl  | 470 | 95 | 12 | US-10-145-639-62  | Sequence 62, Appl  |
| 398 | 6 | 0.8 | 95 | 12 | US-10-140-018-62    | Sequence 62, Appl  | 471 | 95 | 12 | US-10-145-630-62  | Sequence 62, Appl  |
| 399 | 6 | 0.8 | 95 | 12 | US-10-140-021-62    | Sequence 62, Appl  | 472 | 95 | 12 | US-10-145-747-62  | Sequence 62, Appl  |
| 400 | 6 | 0.8 | 95 | 12 | US-10-140-274-62    | Sequence 62, Appl  | 473 | 95 | 12 | US-10-145-752-62  | Sequence 62, Appl  |
| 401 | 6 | 0.8 | 95 | 12 | US-10-140-471-62    | Sequence 62, Appl  | 474 | 95 | 12 | US-10-145-754-62  | Sequence 62, Appl  |
| 402 | 6 | 0.8 | 95 | 12 | US-10-140-807-62    | Sequence 62, Appl  | 475 | 95 | 12 | US-10-145-755-62  | Sequence 62, Appl  |
| 403 | 6 | 0.8 | 95 | 12 | US-10-140-922-62    | Sequence 62, Appl  | 476 | 95 | 12 | US-10-145-818-62  | Sequence 62, Appl  |
| 404 | 6 | 0.8 | 95 | 12 | US-10-140-924-62    | Sequence 62, Appl  | 477 | 95 | 12 | US-10-145-820-62  | Sequence 62, Appl  |
| 405 | 6 | 0.8 | 95 | 12 | US-10-140-926-62    | Sequence 62, Appl  | 478 | 95 | 12 | US-10-145-872-62  | Sequence 62, Appl  |
| 406 | 6 | 0.8 | 95 | 12 | US-10-141-698-62    | Sequence 62, Appl  | 479 | 95 | 12 | US-10-145-873-62  | Sequence 62, Appl  |
| 407 | 6 | 0.8 | 95 | 12 | US-10-141-703-62    | Sequence 62, Appl  | 480 | 95 | 12 | US-10-145-873-62  | Sequence 62, Appl  |
| 408 | 6 | 0.8 | 95 | 12 | US-10-141-704-62    | Sequence 62, Appl  | 481 | 95 | 12 | US-10-145-873-62  | Sequence 62, Appl  |
| 409 | 6 | 0.8 | 95 | 12 | US-10-142-421-62    | Sequence 62, Appl  | 482 | 95 | 12 | US-10-147-482-62  | Sequence 62, Appl  |
| 410 | 6 | 0.8 | 95 | 12 | US-10-142-434-62    | Sequence 62, Appl  | 483 | 95 | 12 | US-10-147-503-62  | Sequence 62, Appl  |
| 411 | 6 | 0.8 | 95 | 12 | US-10-142-767-62    | Sequence 62, Appl  | 484 | 95 | 12 | US-10-147-522-62  | Sequence 62, Appl  |
| 412 | 6 | 0.8 | 95 | 12 | US-10-143-033-62    | Sequence 62, Appl  | 485 | 95 | 12 | US-10-152-401-62  | Sequence 62, Appl  |
| 413 | 6 | 0.8 | 95 | 12 | US-10-144-994-62    | Sequence 62, Appl  | 486 | 95 | 12 | US-10-157-783-62  | Sequence 62, Appl  |
| 414 | 6 | 0.8 | 95 | 12 | US-10-145-628-62    | Sequence 62, Appl  | 487 | 95 | 12 | US-10-158-462-62  | Sequence 62, Appl  |
| 415 | 6 | 0.8 | 95 | 12 | US-10-145-631-62    | Sequence 62, Appl  | 488 | 95 | 12 | US-10-158-792-62  | Sequence 62, Appl  |
| 416 | 6 | 0.8 | 95 | 12 | US-10-145-633-62    | Sequence 62, Appl  | 489 | 95 | 12 | US-10-143-035-62  | Sequence 62, Appl  |
| 417 | 6 | 0.8 | 95 | 12 | US-10-145-746-62    | Sequence 62, Appl  | 490 | 95 | 12 | US-10-143-751-62  | Sequence 62, Appl  |
| 418 | 6 | 0.8 | 95 | 12 | US-10-145-748-62    | Sequence 62, Appl  | 491 | 95 | 12 | US-10-143-822-62  | Sequence 62, Appl  |
| 419 | 6 | 0.8 | 95 | 12 | US-10-145-823-62    | Sequence 62, Appl  | 492 | 95 | 12 | US-10-143-824-62  | Sequence 62, Appl  |
| 420 | 6 | 0.8 | 95 | 12 | US-10-145-826-62    | Sequence 62, Appl  | 493 | 95 | 12 | US-10-143-827-62  | Sequence 62, Appl  |
| 421 | 6 | 0.8 | 95 | 12 | US-10-145-870-62    | Sequence 62, Appl  | 494 | 95 | 12 | US-10-143-869-62  | Sequence 62, Appl  |
| 422 | 6 | 0.8 | 95 | 12 | US-10-145-876-62    | Sequence 62, Appl  | 495 | 95 | 12 | US-10-145-875-62  | Sequence 62, Appl  |
| 423 | 6 | 0.8 | 95 | 12 | US-10-145-959-62    | Sequence 62, Appl  | 496 | 95 | 12 | US-10-145-877-62  | Sequence 62, Appl  |
| 424 | 6 | 0.8 | 95 | 12 | US-10-146-724-62    | Sequence 62, Appl  | 497 | 95 | 12 | US-10-145-958-62  | Sequence 62, Appl  |
| 425 | 6 | 0.8 | 95 | 12 | US-10-146-725-62    | Sequence 62, Appl  | 498 | 95 | 12 | US-10-146-787-62  | Sequence 62, Appl  |
| 426 | 6 | 0.8 | 95 | 12 | US-10-146-793-62    | Sequence 62, Appl  | 499 | 95 | 12 | US-10-146-790-62  | Sequence 62, Appl  |
| 427 | 6 | 0.8 | 95 | 12 | US-10-147-498-62    | Sequence 62, Appl  | 500 | 95 | 12 | US-10-146-793-62  | Sequence 62, Appl  |
| 428 | 6 | 0.8 | 95 | 12 | US-10-147-501-62    | Sequence 62, Appl  | 501 | 95 | 12 | US-10-147-480-62  | Sequence 62, Appl  |
| 429 | 6 | 0.8 | 95 | 12 | US-10-147-504-62    | Sequence 62, Appl  | 502 | 95 | 12 | US-10-147-485-62  | Sequence 62, Appl  |
| 430 | 6 | 0.8 | 95 | 12 | US-10-147-506-62    | Sequence 62, Appl  | 503 | 95 | 12 | US-10-147-486-62  | Sequence 62, Appl  |
| 431 | 6 | 0.8 | 95 | 12 | US-10-147-509-62    | Sequence 62, Appl  | 504 | 95 | 12 | US-10-147-490-62  | Sequence 62, Appl  |
| 432 | 6 | 0.8 | 95 | 12 | US-10-147-510-62    | Sequence 62, Appl  | 505 | 95 | 12 | US-10-147-494-62  | Sequence 62, Appl  |
| 433 | 6 | 0.8 | 95 | 12 | US-10-147-511-62    | Sequence 62, Appl  | 506 | 95 | 12 | US-10-147-498-62  | Sequence 62, Appl  |
| 434 | 6 | 0.8 | 95 | 12 | US-10-147-529-62    | Sequence 62, Appl  | 507 | 95 | 12 | US-10-147-498-62  | Sequence 62, Appl  |
| 435 | 6 | 0.8 | 95 | 12 | US-10-152-397-62    | Sequence 62, Appl  | 508 | 95 | 12 | US-10-147-514-62  | Sequence 62, Appl  |
| 436 | 6 | 0.8 | 95 | 12 | US-10-153-586-62    | Sequence 62, Appl  | 509 | 95 | 12 | US-10-147-524-62  | Sequence 62, Appl  |
| 437 | 6 | 0.8 | 95 | 12 | US-10-158-783-62    | Sequence 62, Appl  | 510 | 95 | 12 | US-10-152-379-62  | Sequence 62, Appl  |
| 438 | 6 | 0.8 | 95 | 12 | US-10-158-786-62    | Sequence 62, Appl  | 511 | 95 | 12 | US-10-152-394-62  | Sequence 62, Appl  |
| 439 | 6 | 0.8 | 95 | 12 | US-10-140-019-62    | Sequence 62, Appl  | 512 | 95 | 12 | US-10-152-406-62  | Sequence 62, Appl  |
| 440 | 6 | 0.8 | 95 | 12 | US-10-140-022-62    | Sequence 62, Appl  | 513 | 95 | 12 | US-10-156-847-62  | Sequence 62, Appl  |
| 441 | 6 | 0.8 | 95 | 12 | US-10-140-861-62    | Sequence 62, Appl  | 514 | 95 | 12 | US-10-157-778-62  | Sequence 62, Appl  |
| 442 | 6 | 0.8 | 95 | 12 | US-10-140-862-62    | Sequence 62, Appl  | 515 | 95 | 12 | US-10-157-793-62  | Sequence 62, Appl  |
| 443 | 6 | 0.8 | 95 | 12 | US-10-141-867-62    | Sequence 62, Appl  | 516 | 95 | 12 | US-10-160-504-62  | Sequence 62, Appl  |
| 444 | 6 | 0.8 | 95 | 12 | US-10-141-700-62    | Sequence 62, Appl  | 517 | 95 | 12 | US-10-145-634-62  | Sequence 62, Appl  |
| 445 | 6 | 0.8 | 95 | 12 | US-10-141-705-62    | Sequence 62, Appl  | 518 | 95 | 12 | US-10-147-520-62  | Sequence 62, Appl  |
| 446 | 6 | 0.8 | 95 | 12 | US-10-141-753-62    | Sequence 62, Appl  | 519 | 95 | 12 | US-10-157-781-62  | Sequence 62, Appl  |
| 447 | 6 | 0.8 | 95 | 12 | US-10-141-758-62    | Sequence 62, Appl  | 520 | 95 | 12 | US-10-176-989-62  | Sequence 62, Appl  |
| 448 | 6 | 0.8 | 95 | 12 | US-10-142-418-62    | Sequence 62, Appl  | 521 | 95 | 12 | US-10-147-491-62  | Sequence 62, Appl  |
| 449 | 6 | 0.8 | 95 | 12 | US-10-142-420-62    | Sequence 62, Appl  | 522 | 95 | 12 | US-10-152-378-62  | Sequence 62, Appl  |
| 450 | 6 | 0.8 | 95 | 12 | US-10-142-422-62    | Sequence 62, Appl  | 523 | 95 | 12 | US-10-152-382-62  | Sequence 62, Appl  |
| 451 | 6 | 0.8 | 95 | 12 | US-10-142-427-62    | Sequence 62, Appl  | 524 | 95 | 12 | US-10-152-383-62  | Sequence 62, Appl  |
| 452 | 6 | 0.8 | 95 | 12 | US-10-142-760-62    | Sequence 62, Appl  | 525 | 95 | 12 | US-10-152-384-62  | Sequence 62, Appl  |
| 453 | 6 | 0.8 | 95 | 12 | US-10-145-821-62    | Sequence 62, Appl  | 526 | 95 | 12 | US-10-152-387-62  | Sequence 62, Appl  |



|     |   |     |    |    |                    |                    |     |
|-----|---|-----|----|----|--------------------|--------------------|-----|
| 673 | 6 | 0.8 | 95 | 15 | US-10-176-921-62   | Sequence 62, Appl  | 746 |
| 674 | 6 | 0.8 | 95 | 15 | US-10-137-865-62   | Sequence 62, Appl  | 747 |
| 675 | 6 | 0.8 | 95 | 15 | US-10-140-474-62   | Sequence 62, Appl  | 748 |
| 676 | 6 | 0.8 | 95 | 15 | US-10-142-431-62   | Sequence 62, Appl  | 749 |
| 677 | 6 | 0.8 | 95 | 15 | US-10-143-114-62   | Sequence 62, Appl  | 750 |
| 678 | 6 | 0.8 | 95 | 15 | US-10-140-002-62   | Sequence 62, Appl  | 751 |
| 679 | 6 | 0.8 | 95 | 15 | US-10-142-419-62   | Sequence 62, Appl  | 752 |
| 680 | 6 | 0.8 | 95 | 15 | US-10-123-243-62   | Sequence 62, Appl  | 753 |
| 681 | 6 | 0.8 | 95 | 15 | US-10-142-443-62   | Sequence 62, Appl  | 754 |
| 682 | 6 | 0.8 | 95 | 15 | US-10-213-181-8    | Sequence 8, Appl   | 755 |
| 683 | 6 | 0.8 | 95 | 15 | US-10-121-050-62   | Sequence 62, Appl  | 756 |
| 684 | 6 | 0.8 | 95 | 15 | US-10-141-755-62   | Sequence 62, Appl  | 757 |
| 685 | 6 | 0.8 | 95 | 15 | US-10-143-032-62   | Sequence 62, Appl  | 758 |
| 686 | 6 | 0.8 | 95 | 15 | US-10-123-108-62   | Sequence 62, Appl  | 759 |
| 687 | 6 | 0.8 | 95 | 15 | US-10-123-216-62   | Sequence 62, Appl  | 760 |
| 688 | 6 | 0.8 | 95 | 15 | US-10-123-281-62   | Sequence 62, Appl  | 761 |
| 689 | 6 | 0.8 | 95 | 15 | US-10-140-921-62   | Sequence 62, Appl  | 762 |
| 690 | 6 | 0.8 | 95 | 15 | US-10-140-928-62   | Sequence 62, Appl  | 763 |
| 691 | 6 | 0.8 | 95 | 15 | US-10-121-045-62   | Sequence 62, Appl  | 764 |
| 692 | 6 | 0.8 | 95 | 15 | US-10-121-292-62   | Sequence 62, Appl  | 765 |
| 693 | 6 | 0.8 | 95 | 15 | US-10-123-903-62   | Sequence 62, Appl  | 766 |
| 694 | 6 | 0.8 | 95 | 15 | US-10-124-819-62   | Sequence 62, Appl  | 767 |
| 695 | 6 | 0.8 | 95 | 15 | US-10-124-822-62   | Sequence 62, Appl  | 768 |
| 696 | 6 | 0.8 | 95 | 15 | US-10-140-925-62   | Sequence 62, Appl  | 769 |
| 697 | 6 | 0.8 | 95 | 15 | US-10-160-498-62   | Sequence 62, Appl  | 770 |
| 698 | 6 | 0.8 | 95 | 15 | US-10-124-824-62   | Sequence 62, Appl  | 771 |
| 699 | 6 | 0.8 | 95 | 15 | US-10-127-825A-62  | Sequence 62, Appl  | 772 |
| 700 | 6 | 0.8 | 95 | 15 | US-10-127-829A-62  | Sequence 62, Appl  | 773 |
| 701 | 6 | 0.8 | 95 | 15 | US-10-127-835A-62  | Sequence 62, Appl  | 774 |
| 702 | 6 | 0.8 | 95 | 15 | US-10-127-839A-62  | Sequence 62, Appl  | 775 |
| 703 | 6 | 0.8 | 95 | 15 | US-10-127-901A-62  | Sequence 62, Appl  | 776 |
| 704 | 6 | 0.8 | 95 | 15 | US-10-128-693A-62  | Sequence 62, Appl  | 777 |
| 705 | 6 | 0.8 | 95 | 15 | US-10-131-813A-62  | Sequence 62, Appl  | 778 |
| 706 | 6 | 0.8 | 95 | 15 | US-10-131-818A-62  | Sequence 62, Appl  | 779 |
| 707 | 6 | 0.8 | 95 | 15 | US-10-131-823A-62  | Sequence 62, Appl  | 780 |
| 708 | 6 | 0.8 | 95 | 15 | US-10-131-824A-62  | Sequence 62, Appl  | 781 |
| 709 | 6 | 0.8 | 95 | 15 | US-10-131-830A-62  | Sequence 62, Appl  | 782 |
| 710 | 6 | 0.8 | 95 | 15 | US-10-131-837A-62  | Sequence 62, Appl  | 783 |
| 711 | 6 | 0.8 | 95 | 15 | US-10-137-872A-62  | Sequence 62, Appl  | 784 |
| 712 | 6 | 0.8 | 95 | 15 | US-10-147-500-62   | Sequence 62, Appl  | 785 |
| 713 | 6 | 0.8 | 95 | 15 | US-10-147-502-62   | Sequence 62, Appl  | 786 |
| 714 | 6 | 0.8 | 95 | 15 | US-10-147-515-62   | Sequence 62, Appl  | 787 |
| 715 | 6 | 0.8 | 95 | 15 | US-10-147-517-62   | Sequence 62, Appl  | 788 |
| 716 | 6 | 0.8 | 95 | 15 | US-10-147-526-62   | Sequence 62, Appl  | 789 |
| 717 | 6 | 0.8 | 95 | 15 | US-10-147-527-62   | Sequence 62, Appl  | 790 |
| 718 | 6 | 0.8 | 95 | 15 | US-10-212-912-8    | Sequence 8, Appl   | 791 |
| 719 | 6 | 0.8 | 95 | 15 | US-10-213-044-8    | Sequence 8, Appl   | 792 |
| 720 | 6 | 0.8 | 95 | 15 | US-10-121-041-62   | Sequence 62, Appl  | 793 |
| 721 | 6 | 0.8 | 95 | 15 | US-10-121-043-62   | Sequence 62, Appl  | 794 |
| 722 | 6 | 0.8 | 95 | 15 | US-10-121-047-62   | Sequence 62, Appl  | 795 |
| 723 | 6 | 0.8 | 95 | 15 | US-10-123-215-62   | Sequence 62, Appl  | 796 |
| 724 | 6 | 0.8 | 95 | 15 | US-10-123-902-62   | Sequence 62, Appl  | 797 |
| 725 | 6 | 0.8 | 95 | 15 | US-10-123-908-62   | Sequence 62, Appl  | 798 |
| 726 | 6 | 0.8 | 95 | 15 | US-10-123-909-62   | Sequence 62, Appl  | 799 |
| 727 | 6 | 0.8 | 95 | 15 | US-10-123-910-62   | Sequence 62, Appl  | 800 |
| 728 | 6 | 0.8 | 95 | 15 | US-10-124-813-62   | Sequence 62, Appl  | 801 |
| 729 | 6 | 0.8 | 95 | 15 | US-10-124-817-62   | Sequence 62, Appl  | 802 |
| 730 | 6 | 0.8 | 95 | 15 | US-10-125-922-62   | Sequence 62, Appl  | 803 |
| 731 | 6 | 0.8 | 95 | 15 | US-10-125-924-62   | Sequence 62, Appl  | 804 |
| 732 | 6 | 0.8 | 95 | 15 | US-10-140-860-62   | Sequence 62, Appl  | 805 |
| 733 | 6 | 0.8 | 95 | 15 | US-10-142-417-62   | Sequence 62, Appl  | 806 |
| 734 | 6 | 0.8 | 95 | 15 | US-10-147-519-62   | Sequence 62, Appl  | 807 |
| 735 | 6 | 0.8 | 95 | 15 | US-10-152-395-62   | Sequence 62, Appl  | 808 |
| 736 | 6 | 0.8 | 95 | 15 | US-10-157-782-62   | Sequence 62, Appl  | 809 |
| 737 | 6 | 0.8 | 95 | 15 | US-10-125-926A-62  | Sequence 62, Appl  | 810 |
| 738 | 6 | 0.8 | 95 | 15 | US-10-125-930A-62  | Sequence 62, Appl  | 811 |
| 739 | 6 | 0.8 | 95 | 15 | US-10-127-831A-62  | Sequence 62, Appl  | 812 |
| 740 | 6 | 0.8 | 95 | 15 | US-10-127-837A-62  | Sequence 62, Appl  | 813 |
| 741 | 6 | 0.8 | 95 | 15 | US-10-127-838B-62  | Sequence 62, Appl  | 814 |
| 742 | 6 | 0.8 | 95 | 15 | US-10-127-842A-62  | Sequence 62, Appl  | 815 |
| 743 | 6 | 0.8 | 95 | 15 | US-10-127-843A-62  | Sequence 62, Appl  | 816 |
| 744 | 6 | 0.8 | 95 | 15 | US-10-127-845A-62  | Sequence 62, Appl  | 817 |
| 745 | 6 | 0.8 | 95 | 15 | US-10-127-846A-62  | Sequence 62, Appl  | 818 |
| 746 | 6 | 0.8 | 95 | 15 | US-10-127-847A-62  | Sequence 62, Appl  | 819 |
| 747 | 6 | 0.8 | 95 | 15 | US-10-127-848A-62  | Sequence 62, Appl  | 820 |
| 748 | 6 | 0.8 | 95 | 15 | US-10-127-849A-62  | Sequence 62, Appl  | 821 |
| 749 | 6 | 0.8 | 95 | 15 | US-10-127-850A-62  | Sequence 62, Appl  | 822 |
| 750 | 6 | 0.8 | 95 | 15 | US-10-127-851A-62  | Sequence 62, Appl  | 823 |
| 751 | 6 | 0.8 | 95 | 15 | US-10-128-684A-62  | Sequence 62, Appl  | 824 |
| 752 | 6 | 0.8 | 95 | 15 | US-10-128-686A-62  | Sequence 62, Appl  | 825 |
| 753 | 6 | 0.8 | 95 | 15 | US-10-128-690A-62  | Sequence 62, Appl  | 826 |
| 754 | 6 | 0.8 | 95 | 15 | US-10-128-691A-62  | Sequence 62, Appl  | 827 |
| 755 | 6 | 0.8 | 95 | 15 | US-10-131-819A-62  | Sequence 62, Appl  | 828 |
| 756 | 6 | 0.8 | 95 | 15 | US-10-131-836A-62  | Sequence 62, Appl  | 829 |
| 757 | 6 | 0.8 | 95 | 15 | US-10-146-729-62   | Sequence 62, Appl  | 830 |
| 758 | 6 | 0.8 | 95 | 15 | US-10-147-484-62   | Sequence 62, Appl  | 831 |
| 759 | 6 | 0.8 | 95 | 15 | US-10-147-508-62   | Sequence 62, Appl  | 832 |
| 760 | 6 | 0.8 | 95 | 15 | US-10-147-512-62   | Sequence 62, Appl  | 833 |
| 761 | 6 | 0.8 | 95 | 15 | US-10-175-735-62   | Sequence 62, Appl  | 834 |
| 762 | 6 | 0.8 | 95 | 15 | US-10-175-736-62   | Sequence 62, Appl  | 835 |
| 763 | 6 | 0.8 | 95 | 15 | US-10-121-040-62   | Sequence 62, Appl  | 836 |
| 764 | 6 | 0.8 | 95 | 15 | US-10-121-056-62   | Sequence 62, Appl  | 837 |
| 765 | 6 | 0.8 | 95 | 15 | US-10-121-061-62   | Sequence 62, Appl  | 838 |
| 766 | 6 | 0.8 | 95 | 15 | US-10-123-235-62   | Sequence 62, Appl  | 839 |
| 767 | 6 | 0.8 | 95 | 15 | US-10-124-818-62   | Sequence 62, Appl  | 840 |
| 768 | 6 | 0.8 | 95 | 15 | US-10-137-868-62   | Sequence 62, Appl  | 841 |
| 769 | 6 | 0.8 | 95 | 15 | US-10-147-492-62   | Sequence 62, Appl  | 842 |
| 770 | 6 | 0.8 | 95 | 15 | US-10-158-782-62   | Sequence 62, Appl  | 843 |
| 771 | 6 | 0.8 | 95 | 15 | US-10-123-905-62   | Sequence 62, Appl  | 844 |
| 772 | 6 | 0.8 | 95 | 15 | US-10-123-907-62   | Sequence 62, Appl  | 845 |
| 773 | 6 | 0.8 | 95 | 15 | US-10-124-815-62   | Sequence 62, Appl  | 846 |
| 774 | 6 | 0.8 | 95 | 15 | US-10-125-928A-62  | Sequence 62, Appl  | 847 |
| 775 | 6 | 0.8 | 95 | 15 | US-10-125-928A-62  | Sequence 62, Appl  | 848 |
| 776 | 6 | 0.8 | 95 | 15 | US-10-127-821A-62  | Sequence 62, Appl  | 849 |
| 777 | 6 | 0.8 | 95 | 15 | US-10-127-822A-62  | Sequence 62, Appl  | 850 |
| 778 | 6 | 0.8 | 95 | 15 | US-10-127-824A-62  | Sequence 62, Appl  | 851 |
| 779 | 6 | 0.8 | 95 | 15 | US-10-127-826A-62  | Sequence 62, Appl  | 852 |
| 780 | 6 | 0.8 | 95 | 15 | US-10-127-827A-62  | Sequence 62, Appl  | 853 |
| 781 | 6 | 0.8 | 95 | 15 | US-10-127-828A-62  | Sequence 62, Appl  | 854 |
| 782 | 6 | 0.8 | 95 | 15 | US-10-127-830A-62  | Sequence 62, Appl  | 855 |
| 783 | 6 | 0.8 | 95 | 15 | US-10-127-832A-62  | Sequence 62, Appl  | 856 |
| 784 | 6 | 0.8 | 95 | 15 | US-10-127-833A-62  | Sequence 62, Appl  | 857 |
| 785 | 6 | 0.8 | 95 | 15 | US-10-127-834A-62  | Sequence 62, Appl  | 858 |
| 786 | 6 | 0.8 | 95 | 15 | US-10-127-836A-62  | Sequence 62, Appl  | 859 |
| 787 | 6 | 0.8 | 95 | 15 | US-10-127-841A-62  | Sequence 62, Appl  | 860 |
| 788 | 6 | 0.8 | 95 | 15 | US-10-127-844A-62  | Sequence 62, Appl  | 861 |
| 789 | 6 | 0.8 | 95 | 15 | US-10-128-687A-62  | Sequence 62, Appl  | 862 |
| 790 | 6 | 0.8 | 95 | 15 | US-10-128-688A-62  | Sequence 62, Appl  | 863 |
| 791 | 6 | 0.8 | 95 | 15 | US-10-128-689A-62  | Sequence 62, Appl  | 864 |
| 792 | 6 | 0.8 | 95 | 15 | US-10-128-694A-62  | Sequence 62, Appl  | 865 |
| 793 | 6 | 0.8 | 95 | 15 | US-10-131-825A-62  | Sequence 8, Appl   | 866 |
| 794 | 6 | 0.8 | 95 | 15 | US-10-213-182-8    | Sequence 8, Appl   | 867 |
| 795 | 6 | 0.8 | 95 | 15 | US-10-230-417-62   | Sequence 62, Appl  | 868 |
| 796 | 6 | 0.8 | 95 | 15 | US-10-131-815A-62  | Sequence 62, Appl  | 869 |
| 797 | 6 | 0.8 | 95 | 15 | US-10-131-817A-62  | Sequence 62, Appl  | 870 |
| 798 | 6 | 0.8 | 95 | 15 | US-10-131-821A-62  | Sequence 62, Appl  | 871 |
| 799 | 6 | 0.8 | 95 | 15 | US-10-131-822A-62  | Sequence 62, Appl  | 872 |
| 800 | 6 | 0.8 | 95 | 15 | US-10-131-828A-62  | Sequence 62, Appl  | 873 |
| 801 | 6 | 0.8 | 95 | 15 | US-10-131-835A-62  | Sequence 62, Appl  | 874 |
| 802 | 6 | 0.8 | 95 | 15 | US-10-137-864A-62  | Sequence 62, Appl  | 875 |
| 803 | 6 | 0.8 | 95 | 15 | US-10-137-869A-62  | Sequence 62, Appl  | 876 |
| 804 | 6 | 0.8 | 95 | 15 | US-10-147-523-62   | Sequence 62, Appl  | 877 |
| 805 | 6 | 0.8 | 95 | 15 | US-10-158-785-62   | Sequence 62, Appl  | 878 |
| 806 | 6 | 0.8 | 95 | 15 | US-10-121-051-62   | Sequence 62, Appl  | 879 |
| 807 | 6 | 0.8 | 95 | 15 | US-10-121-042-62   | Sequence 62, Appl  | 880 |
| 808 | 6 | 0.8 | 95 | 15 | US-10-213-060A-8   | Sequence 8, Appl   | 881 |
| 809 | 6 | 0.8 | 95 | 15 | US-10-123-912-62   | Sequence 62, Appl  | 882 |
| 810 | 6 | 0.8 | 95 | 15 | US-10-223-085-444  | Sequence 244, App  | 883 |
| 811 | 6 | 0.8 | 95 | 15 | US-10-192-007-62   | Sequence 62, Appl  | 884 |
| 812 | 6 | 0.8 | 95 | 15 | US-10-194-359-62   | Sequence 62, Appl  | 885 |
| 813 | 6 | 0.8 | 95 | 15 | US-10-223-084-244  | Sequence 244, App  | 886 |
| 814 | 6 | 0.8 | 95 | 15 | US-10-223-088-244  | Sequence 244, App  | 887 |
| 815 | 6 | 0.8 | 95 | 15 | US-10-223-030-244  | Sequence 244, App  | 888 |
| 816 | 6 | 0.8 | 95 | 15 | US-10-223-087-244  | Sequence 244, App  | 889 |
| 817 | 6 | 0.8 | 95 | 15 | US-10-106-698-696A | Sequence 696A, App | 890 |
| 818 | 6 | 0.8 | 95 | 15 | US-10-127-847A-62  | Sequence 62, Appl  | 891 |
| 819 | 6 | 0.8 | 95 | 15 | US-10-127-848A-62  | Sequence 62, Appl  | 892 |
| 820 | 6 | 0.8 | 95 | 15 | US-10-127-849A-62  | Sequence 62, Appl  | 893 |
| 821 | 6 | 0.8 | 95 | 15 | US-10-127-850A-62  | Sequence 62, Appl  | 894 |
| 822 | 6 | 0.8 | 95 | 15 | US-10-127-851A-62  | Sequence 62, Appl  | 895 |
| 823 | 6 | 0.8 | 95 | 15 | US-10-128-684A-62  | Sequence 62, Appl  | 896 |
| 824 | 6 | 0.8 | 95 | 15 | US-10-128-686A-62  | Sequence 62, Appl  | 897 |
| 825 | 6 | 0.8 | 95 | 15 | US-10-128-690A-62  | Sequence 62, Appl  | 898 |
| 826 | 6 | 0.8 | 95 | 15 | US-10-131-819A-62  | Sequence 62, Appl  | 899 |
| 827 | 6 | 0.8 | 95 | 15 | US-10-131-836A-62  | Sequence 62, Appl  | 900 |
| 828 | 6 | 0.8 | 95 | 15 | US-10-146-729-62   | Sequence           |     |

|     |     |    |                     |                     |     |    |                     |                     |
|-----|-----|----|---------------------|---------------------|-----|----|---------------------|---------------------|
| 819 | 95  | 15 | US-10-223-083-244   | Sequence 244, App   | 892 | 12 | US-10-238-075-1540  | Sequence 1540, App  |
| 820 | 95  | 15 | US-10-175-590-62    | Sequence 62, Appl   | 893 | 12 | US-10-115-223-27    | Sequence 27, Appl   |
| 821 | 95  | 15 | US-10-223-089-244   | Sequence 244, App   | 894 | 12 | US-09-896-522-6     | Sequence 6, Appl    |
| 822 | 95  | 16 | US-10-137-866-62    | Sequence 62, Appl   | 895 | 12 | US-09-893-737-262   | Sequence 262, App   |
| 823 | 95  | 16 | US-10-146-726-62    | Sequence 62, Appl   | 896 | 12 | US-10-325-694-146   | Sequence 146, App   |
| 824 | 95  | 16 | US-10-146-727-62    | Sequence 62, Appl   | 897 | 12 | US-10-325-694-146   | Sequence 151, App   |
| 825 | 95  | 16 | US-10-146-788-62    | Sequence 62, Appl   | 898 | 12 | US-10-106-698-5978  | Sequence 5978, App  |
| 826 | 95  | 16 | US-10-146-788-62    | Sequence 62, Appl   | 899 | 12 | US-09-798-889-59    | Sequence 59, Appl   |
| 827 | 95  | 16 | US-10-152-880-62    | Sequence 62, Appl   | 900 | 12 | US-09-764-877-1531  | Sequence 78, Appl   |
| 828 | 97  | 11 | US-09-969-730-113   | Sequence 113, App   | 901 | 12 | US-09-800-321A-78   | Sequence 1531, App  |
| 829 | 97  | 11 | US-10-203-081-11    | Sequence 11, Appl   | 902 | 12 | US-10-078-929-36    | Sequence 36, Appl   |
| 830 | 97  | 15 | US-10-203-081-16    | Sequence 16, Appl   | 903 | 12 | US-10-169-048-16    | Sequence 16, Appl   |
| 831 | 97  | 15 | US-10-203-081-17    | Sequence 17, Appl   | 904 | 12 | US-10-106-698-5913  | Sequence 6913, App  |
| 832 | 98  | 10 | US-09-894-018-101   | Sequence 101, App   | 905 | 12 | US-09-815-242-4988  | Sequence 4988, App  |
| 833 | 98  | 11 | US-09-774-839-195   | Sequence 195, App   | 906 | 12 | US-09-815-242-10632 | Sequence 10632, App |
| 834 | 98  | 12 | US-10-187-394-1     | Sequence 1, Appl    | 907 | 12 | US-09-738-626-4401  | Sequence 4401, App  |
| 835 | 98  | 12 | US-10-187-394-4     | Sequence 4, Appl    | 908 | 12 | US-09-925-302-750   | Sequence 750, App   |
| 836 | 100 | 9  | US-09-759-143-933   | Sequence 933, App   | 909 | 12 | US-09-764-891-2740  | Sequence 2740, App  |
| 837 | 100 | 9  | US-09-780-669-933   | Sequence 933, App   | 910 | 12 | US-10-101-464A-669  | Sequence 669, App   |
| 838 | 100 | 9  | US-09-822-827-933   | Sequence 933, App   | 911 | 12 | US-10-180-375-133   | Sequence 193, App   |
| 839 | 100 | 10 | US-09-895-793-933   | Sequence 933, App   | 912 | 12 | US-09-791-171-68    | Sequence 68, Appl   |
| 840 | 100 | 10 | US-09-895-814-933   | Sequence 933, App   | 913 | 12 | US-09-804-980-68    | Sequence 8, Appl    |
| 841 | 100 | 12 | US-10-144-678A-933  | Sequence 933, App   | 914 | 12 | US-10-003-278-8     | Sequence 4, Appl    |
| 842 | 100 | 12 | US-10-294-025-933   | Sequence 933, App   | 915 | 12 | US-09-993-756A-4    | Sequence 160, App   |
| 843 | 100 | 14 | US-10-012-896-933   | Sequence 933, App   | 916 | 12 | US-10-166-225A-160  | Sequence 84, Appl   |
| 844 | 102 | 9  | US-09-816-279-2     | Sequence 2, Appl    | 917 | 12 | US-10-062-254-366   | Sequence 366, App   |
| 845 | 102 | 15 | US-10-156-761-8183  | Sequence 8183, App  | 918 | 12 | US-09-877-790-1     | Sequence 1, Appl    |
| 846 | 103 | 12 | US-10-091-007-46    | Sequence 46, App    | 919 | 12 | US-09-764-853-526   | Sequence 526, App   |
| 847 | 103 | 12 | US-10-137-113-28    | Sequence 28, App    | 920 | 12 | US-09-736-457-1672  | Sequence 1672, App  |
| 848 | 103 | 15 | US-10-146-338-9     | Sequence 9, Appl    | 921 | 12 | US-09-903-941-1672  | Sequence 1672, App  |
| 849 | 103 | 15 | US-10-146-338-11    | Sequence 9, Appl    | 922 | 12 | US-09-849-626-1672  | Sequence 1672, App  |
| 850 | 104 | 9  | US-09-925-299-1219  | Sequence 1219, App  | 923 | 12 | US-09-880-505-55    | Sequence 55, Appl   |
| 851 | 104 | 9  | US-09-775-938A-14   | Sequence 14, Appl   | 924 | 12 | US-10-113-872-1672  | Sequence 1672, App  |
| 852 | 104 | 11 | US-09-925-299-1219  | Sequence 1219, App  | 925 | 12 | US-10-051-643-55    | Sequence 55, Appl   |
| 853 | 108 | 10 | US-09-764-877-1708  | Sequence 1708, App  | 926 | 12 | US-10-017-754-1672  | Sequence 1672, App  |
| 854 | 110 | 9  | US-09-733-254-149   | Sequence 149, App   | 927 | 12 | US-10-239-608-5     | Sequence 5, Appl    |
| 855 | 110 | 9  | US-09-904-615-149   | Sequence 149, App   | 928 | 12 | US-10-156-761-8967  | Sequence 8967, App  |
| 856 | 110 | 12 | US-10-055-098-149   | Sequence 149, App   | 929 | 12 | US-10-156-761-12491 | Sequence 12491, App |
| 857 | 110 | 12 | US-10-310-507-3     | Sequence 3, Appl    | 930 | 12 | US-09-946-374-336   | Sequence 336, App   |
| 858 | 110 | 15 | US-10-054-968-149   | Sequence 149, App   | 931 | 12 | US-10-015-387A-336  | Sequence 336, App   |
| 859 | 110 | 15 | US-10-156-761-8628  | Sequence 8628, App  | 932 | 12 | US-10-006-130A-336  | Sequence 336, App   |
| 860 | 110 | 15 | US-10-156-761-13214 | Sequence 13214, App | 933 | 12 | US-10-199-672-360   | Sequence 360, App   |
| 861 | 111 | 9  | US-09-864-761-39500 | Sequence 39500, App | 934 | 12 | US-10-006-172A-336  | Sequence 336, App   |
| 862 | 112 | 10 | US-09-813-271B-2    | Sequence 2, Appl    | 935 | 12 | US-10-187-749-360   | Sequence 360, App   |
| 863 | 112 | 10 | US-09-813-271B-8    | Sequence 8, Appl    | 936 | 12 | US-10-194-457-360   | Sequence 360, App   |
| 864 | 112 | 12 | US-10-342-224-38    | Sequence 38, Appl   | 937 | 12 | US-10-184-642-360   | Sequence 360, App   |
| 865 | 113 | 10 | US-09-764-855-84    | Sequence 84, Appl   | 938 | 12 | US-10-173-691-360   | Sequence 360, App   |
| 866 | 113 | 10 | US-09-813-398-13    | Sequence 13, Appl   | 939 | 12 | US-10-173-692-360   | Sequence 360, App   |
| 867 | 113 | 12 | US-09-890-688-152   | Sequence 152, App   | 940 | 12 | US-10-173-694-360   | Sequence 360, App   |
| 868 | 113 | 15 | US-10-072-349-84    | Sequence 84, Appl   | 941 | 12 | US-10-173-698-360   | Sequence 360, App   |
| 869 | 114 | 10 | US-09-813-459-22    | Sequence 22, Appl   | 942 | 12 | US-10-173-699-360   | Sequence 360, App   |
| 870 | 114 | 14 | US-10-115-406-21    | Sequence 21, Appl   | 943 | 12 | US-10-173-699-360   | Sequence 360, App   |
| 871 | 114 | 15 | US-10-154-333-23    | Sequence 23, Appl   | 944 | 12 | US-10-173-691-360   | Sequence 360, App   |
| 872 | 115 | 10 | US-09-859-211-47    | Sequence 47, Appl   | 945 | 12 | US-10-173-692-360   | Sequence 360, App   |
| 873 | 115 | 10 | US-09-880-708-25    | Sequence 25, Appl   | 946 | 12 | US-10-173-694-360   | Sequence 360, App   |
| 874 | 115 | 11 | US-09-738-626-3792  | Sequence 3792, App  | 947 | 12 | US-10-173-698-360   | Sequence 360, App   |
| 875 | 115 | 11 | US-09-872-856-47    | Sequence 47, Appl   | 948 | 12 | US-10-173-699-360   | Sequence 360, App   |
| 876 | 115 | 15 | US-10-335-483-29    | Sequence 29, Appl   | 949 | 12 | US-10-173-707-360   | Sequence 360, App   |
| 877 | 115 | 16 | US-10-080-170-642   | Sequence 642, App   | 950 | 12 | US-10-174-569-360   | Sequence 360, App   |
| 878 | 116 | 9  | US-09-925-299-1412  | Sequence 1412, App  | 951 | 12 | US-10-174-583-360   | Sequence 360, App   |
| 879 | 116 | 9  | US-09-764-853-763   | Sequence 763, App   | 952 | 12 | US-10-174-587-360   | Sequence 360, App   |
| 880 | 116 | 10 | US-09-896-738-7     | Sequence 7, Appl    | 953 | 12 | US-10-174-589-360   | Sequence 360, App   |
| 881 | 116 | 11 | US-09-925-299-1412  | Sequence 1412, App  | 954 | 12 | US-10-174-591-360   | Sequence 360, App   |
| 882 | 116 | 14 | US-10-115-406-24    | Sequence 24, Appl   | 955 | 12 | US-10-175-736-360   | Sequence 360, App   |
| 883 | 116 | 15 | US-10-154-333-26    | Sequence 26, Appl   | 956 | 12 | US-10-175-742-360   | Sequence 360, App   |
| 884 | 117 | 10 | US-09-736-457-1673  | Sequence 1673, App  | 957 | 12 | US-10-175-744-360   | Sequence 360, App   |
| 885 | 117 | 10 | US-09-802-941-1673  | Sequence 1673, App  | 958 | 12 | US-10-175-745-360   | Sequence 360, App   |
| 886 | 117 | 10 | US-09-849-626-1673  | Sequence 1673, App  | 959 | 12 | US-10-175-748-360   | Sequence 360, App   |
| 887 | 117 | 12 | US-10-113-872-1673  | Sequence 1673, App  | 960 | 12 | US-10-175-751-360   | Sequence 360, App   |
| 888 | 117 | 15 | US-10-017-754-1673  | Sequence 1673, App  | 961 | 12 | US-10-175-754-360   | Sequence 360, App   |
| 889 | 118 | 12 | US-10-029-386-29675 | Sequence 29675, App | 962 | 12 | US-10-176-480-360   | Sequence 360, App   |
| 890 | 120 | 15 | US-10-115-615-24    | Sequence 24, Appl   | 963 | 12 | US-10-176-489-360   | Sequence 360, App   |
| 891 | 121 | 10 | US-09-734-017A-48   | Sequence 48, Appl   | 964 | 12 | US-10-176-754-360   | Sequence 360, App   |

|      |   |     |     |    |                    |                   |
|------|---|-----|-----|----|--------------------|-------------------|
| 965  | 6 | 0.8 | 148 | 12 | US-10-176-755-360  | Sequence 360, App |
| 966  | 6 | 0.8 | 148 | 12 | US-10-176-759-360  | Sequence 360, App |
| 967  | 6 | 0.8 | 148 | 12 | US-10-176-920-360  | Sequence 360, App |
| 968  | 6 | 0.8 | 148 | 12 | US-10-176-922-360  | Sequence 360, App |
| 969  | 6 | 0.8 | 148 | 12 | US-10-176-924-360  | Sequence 360, App |
| 970  | 6 | 0.8 | 148 | 12 | US-10-176-984-360  | Sequence 360, App |
| 971  | 6 | 0.8 | 148 | 12 | US-10-179-508-360  | Sequence 360, App |
| 972  | 6 | 0.8 | 148 | 12 | US-10-179-512-360  | Sequence 360, App |
| 973  | 6 | 0.8 | 148 | 12 | US-10-179-515-360  | Sequence 360, App |
| 974  | 6 | 0.8 | 148 | 12 | US-10-017-306A-336 | Sequence 336, App |
| 975  | 6 | 0.8 | 148 | 12 | US-10-173-702-360  | Sequence 360, App |
| 976  | 6 | 0.8 | 148 | 12 | US-10-173-703-360  | Sequence 360, App |
| 977  | 6 | 0.8 | 148 | 12 | US-10-173-704-360  | Sequence 360, App |
| 978  | 6 | 0.8 | 148 | 12 | US-10-174-574-360  | Sequence 360, App |
| 979  | 6 | 0.8 | 148 | 12 | US-10-176-486-360  | Sequence 360, App |
| 980  | 6 | 0.8 | 148 | 12 | US-10-176-490-360  | Sequence 360, App |
| 981  | 6 | 0.8 | 148 | 12 | US-10-176-752-360  | Sequence 360, App |
| 982  | 6 | 0.8 | 148 | 12 | US-10-176-981-360  | Sequence 360, App |
| 983  | 6 | 0.8 | 148 | 12 | US-10-176-983-360  | Sequence 360, App |
| 984  | 6 | 0.8 | 148 | 12 | US-10-176-988-360  | Sequence 360, App |
| 985  | 6 | 0.8 | 148 | 12 | US-10-179-517-360  | Sequence 360, App |
| 986  | 6 | 0.8 | 148 | 12 | US-10-179-521-360  | Sequence 360, App |
| 987  | 6 | 0.8 | 148 | 12 | US-10-012-064A-336 | Sequence 336, App |
| 988  | 6 | 0.8 | 148 | 12 | US-10-017-867A-336 | Sequence 336, App |
| 989  | 6 | 0.8 | 148 | 12 | US-10-202-475-360  | Sequence 360, App |
| 990  | 6 | 0.8 | 148 | 12 | US-10-012-101B-336 | Sequence 336, App |
| 991  | 6 | 0.8 | 148 | 12 | US-10-012-137A-336 | Sequence 336, App |
| 992  | 6 | 0.8 | 148 | 12 | US-10-012-752A-336 | Sequence 336, App |
| 993  | 6 | 0.8 | 148 | 12 | US-10-012-754A-336 | Sequence 336, App |
| 994  | 6 | 0.8 | 148 | 12 | US-10-013-909A-336 | Sequence 336, App |
| 995  | 6 | 0.8 | 148 | 12 | US-10-013-911A-336 | Sequence 336, App |
| 996  | 6 | 0.8 | 148 | 12 | US-10-013-912A-336 | Sequence 336, App |
| 997  | 6 | 0.8 | 148 | 12 | US-10-015-610A-336 | Sequence 336, App |
| 998  | 6 | 0.8 | 148 | 12 | US-10-015-653A-336 | Sequence 336, App |
| 999  | 6 | 0.8 | 148 | 12 | US-10-015-671A-336 | Sequence 336, App |
| 1000 | 6 | 0.8 | 148 | 15 | US-10-187-755-360  | Sequence 360, App |

ALIGNMENTS

|  |     |   |  |  |  |  |  |  |  |
|--|-----|---|--|--|--|--|--|--|--|
| RESULT 1   |     |   |  |  |  |  |  |  |  |
| US-10-320-800-10   |     |   |  |  |  |  |  |  |  |
| ; Sequence 10, Application US/10320800                       |     |   |  |  |  |  |  |  |  |
| ; Publication No. US20030215469A1                            |     |   |  |  |  |  |  |  |  |
| ; GENERAL INFORMATION:                                       |     |   |  |  |  |  |  |  |  |
| ; APPLICANT: ROBINSON, ANDREW                                |     |   |  |  |  |  |  |  |  |
| ; APPLICANT: GORRINGE, ANDREW                                |     |   |  |  |  |  |  |  |  |
| ; APPLICANT: HUDSON, MICHAEL                                 |     |   |  |  |  |  |  |  |  |
| ; APPLICANT: REDDIN, KAREN                                   |     |   |  |  |  |  |  |  |  |
| ; TITLE OF INVENTION: MULTICOMPONENT MENINGOCOCCAL VACCINE   |     |   |  |  |  |  |  |  |  |
| ; FILE REFERENCE: 1591.0790001                               |     |   |  |  |  |  |  |  |  |
| ; CURRENT APPLICATION NUMBER: US/10/320,800                  |     |   |  |  |  |  |  |  |  |
| ; CURRENT FILING DATE: 2002-12-17                            |     |   |  |  |  |  |  |  |  |
| ; PRIOR APPLICATION NUMBER: PCT/GB99/03626                   |     |   |  |  |  |  |  |  |  |
| ; PRIOR FILING DATE: 1999-11-02                              |     |   |  |  |  |  |  |  |  |
| ; NUMBER OF SEQ ID NOS: 75                                   |     |   |  |  |  |  |  |  |  |
| ; SOFTWARE: PatentIn version 3.1                             |     |   |  |  |  |  |  |  |  |
| ; SEQ ID NO 10   |     |   |  |  |  |  |  |  |  |
| ; LENGTH: 769  |     |   |  |  |  |  |  |  |  |
| ; TYPE: PAT  |     |   |  |  |  |  |  |  |  |
| ; ORGANISM: Neisseria meningitidis                           |     |   |  |  |  |  |  |  |  |
| US-10-320-800-10   |     |   |  |  |  |  |  |  |  |
| Query Match 46.3%; Score 356; DB 12; Length 769;             |     |   |  |  |  |  |  |  |  |
| Best Local Similarity 99.8%;                                 |     |   |  |  |  |  |  |  |  |
| Matches 556; Conservative 0; Mismatches 2; Indels 0; Gaps 0; |     |   |  |  |  |  |  |  |  |
| QY   | 212 | AAPAKQAAAPAKOTNIDFRKDGKNGAGIIEALGAFAGQFDISQQHDIHIVTLKXHTLPT 271 |  |  |  |  |  |  |  |
| Db   | 212 | AAPAKQAAAPAKOTNIDFRKDGKNGAGIIEALGAFAGQFDISQQHDIHIVTLKXHTLPT 271 |  |  |  |  |  |  |  |
| QY   | 272 | TLORSLDVADFKTPVQKVLTKRLNNDTQLIITAGNWLNVKSAAPGYFTFOVLPKKQNL 331  |  |  |  |  |  |  |  |

|    |     |   |  |  |  |  |  |  |  |
|----|-----|---|--|--|--|--|--|--|--|
| Db | 272 | TLORSLDVADFKTPVQKVLTKRLNNDTQLIITAGNWLNVKSAAPGYFTFOVLPKKQNL 331    |  |  |  |  |  |  |  |
| QY | 332 | ESGVNNAKPTFTGRKISLDFQDVEIRTIQLILAKESGMNIVASDSVNGKMTLSLKDVPW 391   |  |  |  |  |  |  |  |
| Db | 332 | ESGVNNAKPTFTGRKISLDFQDVEIRTIQLILAKESGMNIVASDSVNGKMTLSLKDVPW 391   |  |  |  |  |  |  |  |
| QY | 392 | DQALDLVMOARNLDMQOQGNIVNIAPRDELLAKDKAFLQAEKDIALGALYSQNFOLKYK 451   |  |  |  |  |  |  |  |
| Db | 392 | DQALDLVMOARNLDMQOQGNIVNIAPRDELLAKDKAFLQAEKDIALGALYSQNFOLKYK 451   |  |  |  |  |  |  |  |
| QY | 452 | NVEEFSIRLNDADTTGNRNTLVSGRGSVLIDPATNTLIIVTDTSRVIEKFKRLIDELDV 511   |  |  |  |  |  |  |  |
| Db | 452 | NVEEFSIRLNDADTTGNRNTLVSGRGSVLIDPATNTLIIVTDTSRVIEKFKRLIDELDV 511   |  |  |  |  |  |  |  |
| QY | 512 | PAQVMIEARIVAAAGFSRDLGVKFGATKKLKNDDTSAFGHWGVSFGGDDKWAETK 571       |  |  |  |  |  |  |  |
| Db | 512 | PAQVMIEARIVAAAGFSRDLGVKFGATKKLKNDDTSAFGHWGVSFGGDDKWAETK 571       |  |  |  |  |  |  |  |
| QY | 572 | INLPITAAANSISLVRAISSGALNLELSASELSKTKTLANPRVLTONRKEAKIESGYEI 631   |  |  |  |  |  |  |  |
| Db | 572 | INLPITAAANSISLVRAISSGALNLELSASELSKTKTLANPRVLTONRKEAKIESGYEI 631   |  |  |  |  |  |  |  |
| QY | 632 | PFTVTSIANGSGSSTNTELKKAIVLGLTVTPNITPDGQIIMTVKINKDSPACASGNQITLC 691 |  |  |  |  |  |  |  |
| Db | 632 | PFTVTSIANGSGSSTNTELKKAIVLGLTVTPNITPDGQIIMTVKINKDSPACASGNQITLC 691 |  |  |  |  |  |  |  |
| QY | 692 | ISTKNLNTQAWVNGGTLIVGGIYEEDNGNTLTKVPLGDIPIVIGNLFKTRGKKTDRREL 751   |  |  |  |  |  |  |  |
| Db | 692 | ISTKNLNTQAWVNGGTLIVGGIYEEDNGNTLTKVPLGDIPIVIGNLFKTRGKKTDRREL 751   |  |  |  |  |  |  |  |
| QY | 752 | LIFITPRIMTAGNSLRY 769   |  |  |  |  |  |  |  |
| Db | 752 | LIFITPRIMTAGNSLRY 769   |  |  |  |  |  |  |  |

RESULT 2

US-09-815-242-10361  
; Sequence 10361, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zykkind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10361  
; LENGTH: 654  
; TYPE: PAT

```
; ORGANISM: Escherichia coli
US-09-815-242-10361

Query Match
Best Local Similarity 1.2%; Score 9; DB 9; Length 654;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDIP 733
DB 563 KVPLLGDIP 571
|||||

RESULT 3
US-10-156-761-11589
; Sequence 11589, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11589
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11589

Query Match
Best Local Similarity 1.0%; Score 8; DB 15; Length 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 KAAPAAPA 147
DB 14 KAAPAAPA 21
|||||

RESULT 4
US-09-815-242-5092
; Sequence 5092, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578

; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5092

Query Match
Best Local Similarity 1.0%; Score 8; DB 9; Length 776;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 726 VPLLGDIP 733
DB 562 VPLLGDIP 569
|||||

RESULT 5
US-10-211-962-49
; Sequence 49, Application US/10211962
; Publication No. US20030082640A1
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/10/211,962
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: US/09/562,737
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 49
; LENGTH: 1024
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-10-211-962-49

Query Match
Best Local Similarity 1.0%; Score 8; DB 15; Length 1024;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 QAAASAKQ 201
DB 923 QAAASAKQ 930
|||||

RESULT 6
US-09-912-020-302
; Sequence 302, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001DVI
; FILE REFERENCE: ELITRA.001DVI
```

```
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 302
; LENGTH: 2383
; TYPE: PRT
; ORGANISM: E. Coli
US-09-912-020-302

Query Match      1.0%; Score 8; DB 9; Length 2383;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      507 DELDVPAQ 514
DB      138 DELDVPAQ 145

RESULT 7
US-10-006-869-2429
; Sequence 2429, Application US/10006869
; Publication No. US20030082166A1
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Symonds, James Matthew
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
; FILE REFERENCE: 100086.407C7
; CURRENT APPLICATION NUMBER: US/10/006,869
; CURRENT FILING DATE: 2001-12-03
; NUMBER OF SEQ ID NOS: 4052
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2429
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Representative cyclic modulating agent based on
; OTHER INFORMATION: PB-cadherin cell adhesion recognition sequence
US-10-006-869-2429

Query Match      0.9%; Score 7; DB 15; Length 7;
Best Local Similarity 100.0%; Pred. No. 6.1e+05;
Matches      7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      504 KLIDELD 510
DB      1 KLIDELD 7

RESULT 8
US-09-572-404B-2232
; Sequence 2232, Application US/09572404B
; Publication No. US20030078374A1
; GENERAL INFORMATION:
; APPLICANT: Proteom Ltd
; TITLE OF INVENTION: Complementary peptide ligands from the human genome
; FILE REFERENCE: Human patent
; CURRENT APPLICATION NUMBER: US/09/572,404B
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 4203
; SOFTWARE: ProtPatent version 1.0
; SEQ ID NO 2232
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Homo Sapiens
; FEATURE:

; OTHER INFORMATION: sequence located in H1F4 at 3-12 and may interact with Sequence
; OTHER INFORMATION: this patent.
US-09-572-404B-2232

Query Match      0.9%; Score 7; DB 11; Length 10;
Best Local Similarity 100.0%; Pred. No. 21;
Matches      7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      141 AAPAAPA 147
DB      3 AAPAAPA 9

RESULT 9
US-09-572-404B-2234
; Sequence 2234, Application US/09572404B
; Publication No. US20030078374A1
; GENERAL INFORMATION:
; APPLICANT: Proteom Ltd
; TITLE OF INVENTION: Complementary peptide ligands from the human genome
; FILE REFERENCE: Human patent
; CURRENT APPLICATION NUMBER: US/09/572,404B
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 4203
; SOFTWARE: ProtPatent version 1.0
; SEQ ID NO 2234
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Homo Sapiens
; FEATURE:
; OTHER INFORMATION: sequence located in H1F4 at 4-13 and may interact with Sequence
; OTHER INFORMATION: this patent.
US-09-572-404B-2234

Query Match      0.9%; Score 7; DB 11; Length 10;
Best Local Similarity 100.0%; Pred. No. 21;
Matches      7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      141 AAPAAPA 147
DB      4 AAPAAPA 10

RESULT 10
US-09-572-404B-2416
; Sequence 2416, Application US/09572404B
; Publication No. US20030078374A1
; GENERAL INFORMATION:
; APPLICANT: Proteom Ltd
; TITLE OF INVENTION: Complementary peptide ligands from the human genome
; FILE REFERENCE: Human patent
; CURRENT APPLICATION NUMBER: US/09/572,404B
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 4203
; SOFTWARE: ProtPatent version 1.0
; SEQ ID NO 2416
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Homo Sapiens
; FEATURE:
; OTHER INFORMATION: sequence located in H1F4 at 4-13 and may interact with Sequence
; OTHER INFORMATION: this patent.
US-09-572-404B-2416

Query Match      0.9%; Score 7; DB 11; Length 10;
Best Local Similarity 100.0%; Pred. No. 21;
Matches      7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      141 AAPAAPA 147
DB      3 AAPAAPA 9
```

```
RESULT 11
US-09-572-404B-2418
; Sequence 2418, Application US/09572404B
; Publication No. US20030078374A1
; GENERAL INFORMATION:
; APPLICANT: Proteom Ltd
; TITLE OF INVENTION: Complementary peptide ligands from the human genome
; FILE REFERENCE: Human patent
; CURRENT APPLICATION NUMBER: US/09/572.404B
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 4203
; SOFTWARE: ProPatent version 1.0
; SEQ ID NO 2418
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: sequence located in H1F4 at 3-12 and may interact with Sequence 2
; OTHER INFORMATION: this patent.
US-09-572-404B-2418

Query Match          0.9%; Score 7; DB 11; Length 10;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPA 147
Db 4 AAPAAPA 10

RESULT 12
US-10-225-567A-1405
; Sequence 1405, Application US/10225567A
; Publication No. US20030113798A1
; GENERAL INFORMATION:
; APPLICANT: Lifespan Biosciences
; APPLICANT: Brown, Joseph P.
; APPLICANT: Burner, Glena C.
; APPLICANT: Roush, Christine L.
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 1920-4-4
; CURRENT APPLICATION NUMBER: US/10/225.567A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/257,144
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 2292
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1405
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-225-567A-1405

Query Match          0.9%; Score 7; DB 15; Length 19;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 NLESGGV 336
Db 7 NLESGGV 13

RESULT 13
US-10-300-083-36
; Sequence 36, Application US/10300083
; Publication No. US20030153502A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA
; TITLE OF INVENTION: SYNTHETIC APPROACH TO DESIGNED CHEMICAL
; STRUCTURES
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MUEITING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.
```

```
STREET: 119 No. US20030153502A1ath Fourth Street, Suite 203
CITY: Minneapolis
STATE: Minnesota
COUNTRY: U.S.A.
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/10/300.083
APPLICATION NUMBER: US/10/300.083
FILING DATE: 20-NOV-2003
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/194.296
FILING DATE: 15-Oct-1999
APPLICATION NUMBER: US 08/653,632
FILING DATE: 24-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: MCCORMACK, MYRA M.
REGISTRATION NUMBER: 36,602
REFERENCE/DOCKET NUMBER: 110.00330220
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1225
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-10-300-083-36

Query Match          0.9%; Score 7; DB 12; Length 33;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 345 GRKISLD 351
Db 27 GRKISLD 33

RESULT 14
US-09-851-873-47
; Sequence 47, Application US/09851873
; Publication No. US20030165488A1
; GENERAL INFORMATION:
; APPLICANT: Kletzien, Rolf F
; APPLICANT: Reardon, Ilene M
; APPLICANT: Weiland, Katherine L
; TITLE OF INVENTION: HUMAN CASPASE-12 MATERIALS AND METHODS
; FILE REFERENCE: 28341/00233
; CURRENT APPLICATION NUMBER: US/09/851,873
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: peptide
US-09-851-873-47

Query Match          0.9%; Score 7; DB 12; Length 51;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 KOKIVKV 44
```



Db 6 KQKIVK 12

RESULT 15

US-09-764-891-5018

; Sequence 5018, Application US/09764891

; Publication No. US20030077808A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC006

; CURRENT APPLICATION NUMBER: US/09/764,891

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 10231

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5018

; LENGTH: 56

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-764-891-5018

Query Match 0.9%; Score 7; DB 11; Length 56;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 KIISGLF 13

Db 31 KIISGLF 37

RESULT 16

US-09-764-891-5018

; Sequence 5018, Application US/09764891

; Publication No. US20030077808A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC006

; CURRENT APPLICATION NUMBER: US/09/764,891

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 10231

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5018

; LENGTH: 56

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-764-891-5018

Query Match 0.9%; Score 7; DB 11; Length 56;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 KIISGLF 13

Db 31 KIISGLF 37

RESULT 17

US-10-029-386-30401

; Sequence 30401, Application US/10029386

; Publication No. US20030194704A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Hanzel, David K.

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO

; FILE REFERENCE: AEOMICA-X-2

; CURRENT APPLICATION NUMBER: US/10/029,386

; CURRENT FILING DATE: 2001-12-20

; NUMBER OF SEQ ID NOS: 34288

; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 30401

; LENGTH: 67

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: MAP TO CHR6.1

; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 24

; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 13

; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 11

; OTHER INFORMATION: SWISSPROT HIT: P35979, EVALUATE 1.00e-31

US-10-029-386-30401

Query Match 0.9%; Score 7; DB 12; Length 67;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 412 IVNIAPR 418

Db 11 IVNIAPR 17

RESULT 18

US-09-943-123-1

; Sequence 1, Application US/09943123

; Publication No. US20020182701A1

; GENERAL INFORMATION:

; APPLICANT: CHANG, Y-H

; APPLICANT: VETRO, J.A.

; TITLE OF INVENTION: Dominant Negative Variants of Methionine Aminopeptidase

; TITLE OF INVENTION: 2 ("MetAp2") and Clinical Uses Therefor

; FILE REFERENCE: 16153-8007

; CURRENT APPLICATION NUMBER: US/09/943,123

; CURRENT FILING DATE: 2001-08-30

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 71

; TYPE: PRT

; ORGANISM: Human polylysine

US-09-943-123-1

Query Match 0.9%; Score 7; DB 10; Length 71;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 538 GATGKKK 544

Db 59 GATGKKK 65

RESULT 19

```

US-09-943-123-2
; Sequence 2, Application US/09943123
; Publication No. US20020182701A1
; GENERAL INFORMATION:
; APPLICANT: CHANG, Y-H
; APPLICANT: VETRO, J.A.
; APPLICANT: MICKA, W.S.
; TITLE OF INVENTION: Dominant Negative Variants of Methionine Aminopeptidase
; TITLE OF INVENTION: 2 ("MetAP2") and Clinical Uses Therefor
; FILE REFERENCE: 16153-8007
; CURRENT APPLICATION NUMBER: US/09/943,123
; CURRENT FILING DATE: 2001-08-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 2
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Mouse polyllysine
US-09-943-123-2
Query Match 0.9%; Score 7; DB 10; Length 71;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 538 GATGKKK 544
Db 59 GATGKKK 55

RESULT 20
US-09-820-843A-89
; Sequence 89, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 89
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: AE004587.5 hypothetical protein
; NAME/KEY: misc.feature
; OTHER INFORMATION: Gi|9947556
US-09-820-843A-89
Query Match 0.9%; Score 7; DB 11; Length 78;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 141 AAPAAPA 147
Db 29 AAPAAPA 35

RESULT 21
US-09-738-626-6042
; Sequence 6042, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO

```

```

; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6042
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6042
Query Match 0.9%; Score 7; DB 10; Length 94;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 706 GGTLLVG 712
Db 31 GGTLLVG 37

RESULT 22
US-10-072-159-7
; Sequence 7, Application US/10072159
; Publication No. US20020151498A1
; GENERAL INFORMATION:
; APPLICANT: Bihain, Bernard
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Yen-Potin, Frances
; TITLE OF INVENTION: Lipoprotein-regulating medicaments
; FILE REFERENCE: GENSET.036APC
; CURRENT APPLICATION NUMBER: US/10/072,159
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: 09/485,316
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: FR 98/05032
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: PCT IB98/01256
; PRIOR FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent.pm
; SEQ ID NO 7
; LENGTH: 129
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: 1..129
; OTHER INFORMATION: fragment 117..245 of ref swissprot P02745
US-10-072-159-7
Query Match 0.9%; Score 7; DB 14; Length 129;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 319 YFTFQVL 325
Db 44 YFTFQVL 50

RESULT 23
US-10-060-845-7
; Sequence 7, Application US/10060845

```

Publication No. US20020165154A1  
; GENERAL INFORMATION:  
; APPLICANT: Bihain, Bernard  
; APPLICANT: Bougueleret, Lydie  
; APPLICANT: Yen-Pocin, Frances  
; TITLE OF INVENTION: Lipoprotein-regulating medicaments  
; FILE REFERENCE: GENSET.036APC  
; CURRENT APPLICATION NUMBER: US/10/060,845  
; CURRENT FILING DATE: 2002-01-29  
; PRIOR APPLICATION NUMBER: U.S. 09/485,316  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: FR 97/10088  
; PRIOR FILING DATE: 1997-08-06  
; PRIOR APPLICATION NUMBER: FR 98/05032  
; PRIOR FILING DATE: 1998-04-22  
; PRIOR APPLICATION NUMBER: PCT IB98/01256  
; PRIOR FILING DATE: 1998-08-06  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patent.pm  
; SEQ ID NO 7  
; LENGTH: 129  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIMILAR  
; LOCATION: 1..129  
; OTHER INFORMATION: fragment 117..245 of ref swissprot P02745  
US-10-060-845-7

Query Match 0.9%; Score 7; DB 14; Length 129;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 319 YTFQVL 325  
Db 44 YTFQVL 50  
|||||

RESULT 24  
US-09-893-737-144  
; Sequence 144, Application US/09893737  
; Patent No. US20020110855A1  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; APPLICANT: Presnell, Scott R.  
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS  
; FILE REFERENCE: 00-41  
; CURRENT APPLICATION NUMBER: US/09/893,737  
; CURRENT FILING DATE: 2001-06-28  
; PRIOR APPLICATION NUMBER: US 60/215,446  
; PRIOR FILING DATE: 2000-06-30  
; NUMBER OF SEQ ID NOS: 329  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 144  
; LENGTH: 130  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-893-737-144

Query Match 0.9%; Score 7; DB 10; Length 130;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 SSRARLV 102  
Db 47 SSRARLV 53  
|||||

RESULT 25  
US-10-156-761-13774  
; Sequence 13774, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: Van Rooijen, Gijls

APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 13774  
; LENGTH: 157  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-13774

Query Match 0.9%; Score 7; DB 15; Length 157;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 LLAOKKA 427  
Db 110 LLAOKKA 116  
|||||

RESULT 26  
US-10-156-761-10017  
; Sequence 10017, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 10017  
; LENGTH: 158  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-10017

Query Match 0.9%; Score 7; DB 15; Length 158;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPA 147  
Db 12 AAPAAPA 18  
|||||

RESULT 27  
US-10-032-201B-153  
; Sequence 153, Application US/10032201B  
; Publication No. US20030167524A1  
; GENERAL INFORMATION:  
; APPLICANT: Van Rooijen, Gijls

```
; APPLICANT: Deckers, Harm
; APPLICANT: Heifetz, Peter Bernard
; APPLICANT: Briggs, Steven
; APPLICANT: Delmia, Bipin Kumar
; APPLICANT: Del Val, Greg
; APPLICANT: Zaplachinski, Steve
; APPLICANT: Moloney, Maurice
; TITLE OF INVENTION: METHODS FOR THE PRODUCTION OF MULTIMERIC PROTEINS, AND RELATED
; FILE REFERENCE: 38614 351B
; CURRENT APPLICATION NUMBER: US/10/032,201B
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 313
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 153
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-032-201B-153

Query Match          0.9%; Score 7; DB 12; Length 166;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 655 GLTVTPN 661
Db 41 GLTVTPN 47

RESULT 28
US-10-102-806-656
; Sequence 656, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1C1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 656
; LENGTH: 167
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (59)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (73)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-806-656

Query Match          0.9%; Score 7; DB 15; Length 167;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 SPAKQQA 195
Db 117 SPAKQQA 123

RESULT 29
US-10-050-704-331
; Sequence 331, Application US/10050704
; Publication No. US20030050442A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 62 Human Secreted Proteins
; FILE REFERENCE: P2039P1
; CURRENT APPLICATION NUMBER: US/10/050,704
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: 09/684,524
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: PCT/US00/08979
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/128,693
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 60/130,991
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 344
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 331
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (22)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-050-704-331

Query Match          0.9%; Score 7; DB 15; Length 182;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 598 LSASESL 604
Db 117 LSASESL 123

RESULT 30
US-10-156-761-9729
; Sequence 9729, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9729
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9729

Query Match          0.9%; Score 7; DB 15; Length 194;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPA 147
Db 7 AAPAAPA 13

RESULT 31
US-09-815-242-5457
```

```
; Sequence 5457, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5457
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5457

Query Match          0.9%; Score 7; DB 9; Length 197;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      737 NLFKTRG 743
Db      164 NLFKTRG 170

RESULT 32
US-09-815-242-12680
; Sequence 12680, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5457
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5457

Query Match          0.9%; Score 7; DB 9; Length 197;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      737 NLFKTRG 743
Db      164 NLFKTRG 170

RESULT 33
US-09-764-898-187
; Sequence 187, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P0201
; CURRENT APPLICATION NUMBER: US/09/764,898
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 311
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 187
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-898-187

Query Match          0.9%; Score 7; DB 9; Length 205;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      139 VKAAPAA 145
Db      154 VKAAPAA 160

RESULT 34
US-09-738-626-3591
; Sequence 3591, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
```

```

; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3591
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3591

Query Match      0.9%; Score 7; DB 10; Length 205;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPA 147
Db 131 AAPAAPA 137

RESULT 35
US-10-156-761-13477
; Sequence 13477, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13477
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13477

Query Match      0.9%; Score 7; DB 15; Length 208;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 645 TNETLKK 651
Db 131 TNETLKK 137

RESULT 36
US-09-811-284-177
; Sequence 177, Application US/09811284
; Patent No. US20020058306A1
; GENERAL INFORMATION:
; APPLICANT: Vogeli, Gabriel
; TITLE OF INVENTION: NO. US20020058306A1e1 G Protein-Coupled Receptors
; FILE REFERENCE: 00167US1
; CURRENT APPLICATION NUMBER: US/09/811,284
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/189,783
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/189,907
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/189,918
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/189,960

; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/189,917
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/192,945
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/192,916
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/192,923
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/192,933
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/192,830
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/192,234
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/192,155
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/192,935
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 258
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 177
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-811-284-177

Query Match      0.9%; Score 7; DB 9; Length 217;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 598 LSASESL 604
Db 130 LSASESL 136

RESULT 37
US-10-229-567-4
; Sequence 4, Application US/10229567
; Publication No. US20030092080A1
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Cohavy, Offer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; Ulcerative Colitis, and Clinical Subtypes Thereof, Using
; Microbial UC pANCA antigens

; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/229,567
; FILING DATE: 27-Aug-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/417,264
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 09/041,889
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
; TELECOMMUNICATION INFORMATION:
```

TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 218 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..218  
OTHER INFORMATION: /note= "product = Human Histone  
H1-S-4"  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-10-229-567-4

Query Match 0.9%; Score 7; DB 15; Length 218;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPA 147  
DB 6 AAPAAPA 12

RESULT 38  
US-10-128-714-3122  
; Sequence 3122, Application US/10128714  
; Publication No. US20030119013A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Bo  
; APPLICANT: Hu, Weng  
; APPLICANT: Tishkoff, Daniel  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Eroskin, Alexey M  
; APPLICANT: Lemieux, Sebastien M  
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
; TITLE OF INVENTION: Methods of Use  
; FILE REFERENCE: 10182-018-999  
; CURRENT APPLICATION NUMBER: US/10/128,714  
; CURRENT FILING DATE: 2002-04-23  
; PRIOR APPLICATION NUMBER: US 60/285,697  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/287,066  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/295,890  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/303,899  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/316,362  
; PRIOR FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 8603  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3122  
; LENGTH: 221  
; TYPE: PRT  
; ORGANISM: Aspergillus fumigatus  
US-10-128-714-3122

Query Match 0.9%; Score 7; DB 15; Length 221;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 419 DELLAKD 425  
DB 50 DELLAKD 56

RESULT 39  
US-09-970-711-11  
; Sequence 11, Application US/09970711  
; Patent No. US20020081279A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Adam

APPLICANT: Cotten, Matthew  
APPLICANT: Chiocca, Susanna  
APPLICANT: Kurzbaumer, Robert  
APPLICANT: Schaffner, Gottfried  
TITLE OF INVENTION: Chicken Embryo Lethal Orphan (CELO) Virus  
FILE REFERENCE: 0652.1800001  
CURRENT APPLICATION NUMBER: US/09/970,711  
CURRENT FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: 09/171,461  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: PCT/EP97/01944  
PRIOR FILING DATE: 1997-04-18  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 11  
LENGTH: 223  
TYPE: PRT  
ORGANISM: CELO Virus  
FEATURE:  
OTHER INFORMATION: Position: 17559..18230 /gene: L3 /product: L3 pVI  
US-09-970-711-11

Query Match 0.9%; Score 7; DB 9; Length 223;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPA 147  
DB 139 AAPAAPA 145

RESULT 40  
US-09-800-729-117  
; Sequence 117, Application US/09800729  
; Patent No. US20020068319A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 32 Human secreted proteins  
; FILE REFERENCE: P204421  
; CURRENT APPLICATION NUMBER: US/09/800,729  
; CURRENT FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: PCT/US00/26013  
; PRIOR FILING DATE: 2000-09-22  
; PRIOR APPLICATION NUMBER: 60/155,709  
; PRIOR FILING DATE: 1999-09-24  
; NUMBER OF SEQ ID NOS: 217  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 117  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (47)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (49)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-800-729-117

Query Match 0.9%; Score 7; DB 9; Length 229;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 319 YFTFQVL 325  
DB 144 YFTFQVL 150

Search completed: December 9, 2003, 10:41:16  
Job time : 50 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 10:33:03 ; Search time 21 Seconds

(without alignments)  
3521.506 Million cell updates/sec

Title: US-09-701-271A-2

Perfect score: 769

Sequence: 1 MNTKTKIISGLFVATAAQ.....ELLIFITPRIMGTAGNSLRY 769

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR.76.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 548   | 71.3        | 761    | 2 A81985 | pilus secretin NMA |
| 2          | 150   | 19.5        | 720    | 2 S70838 | pilQ protein - Nei |
| 3          | 142   | 18.5        | 711    | 2 A37051 | outer membrane pro |
| 4          | 12    | 1.6         | 660    | 2 B34459 | pullulanase secret |
| 5          | 11    | 1.4         | 445    | 2 H64067 | hypothetical prote |
| 6          | 11    | 1.4         | 445    | 2 S37345 | pilQ protein - Pse |
| 7          | 11    | 1.4         | 714    | 2 A83016 | type 4 fimbrial bi |
| 8          | 10    | 1.3         | 271    | 2 S22668 | hypothetical prote |
| 9          | 10    | 1.3         | 642    | 2 T42133 | type II secretion  |
| 10         | 10    | 1.3         | 649    | 2 S2858  | outD protein - Erw |
| 11         | 10    | 1.3         | 678    | 2 S46963 | exed protein - Aer |
| 12         | 10    | 1.3         | 909    | 2 H87729 | protein Y23HSA.7 [ |
| 13         | 9     | 1.2         | 607    | 2 C40361 | virC-region hypoth |
| 14         | 9     | 1.2         | 654    | 2 H65125 | type III secretion |
| 15         | 9     | 1.2         | 654    | 2 S36653 | probable general s |
| 16         | 9     | 1.2         | 658    | 2 S26653 | xcpQ protein - Pse |
| 17         | 9     | 1.2         | 705    | 2 F70352 | conserved hypotet  |
| 18         | 9     | 1.2         | 710    | 2 S28014 | outD protein - Erw |
| 19         | 9     | 1.2         | 712    | 2 B47021 | pectic enzyme secr |
| 20         | 9     | 1.2         | 748    | 2 A45243 | envelope protein H |
| 21         | 8     | 1.0         | 111    | 2 G70772 | hypothetical prote |
| 22         | 8     | 1.0         | 231    | 2 A81036 | mannose-1-phosphat |
| 23         | 8     | 1.0         | 231    | 2 C81981 | probable sugar-pho |
| 24         | 8     | 1.0         | 306    | 2 A83526 | homoserine kinase  |
| 25         | 8     | 1.0         | 311    | 2 E81703 | phospholipase D fa |
| 26         | 8     | 1.0         | 329    | 2 S62477 | probable serine/th |
| 27         | 8     | 1.0         | 401    | 2 S16738 | porphobilinogen sy |
| 28         | 8     | 1.0         | 412    | 2 B65134 | protein transport  |
| 29         | 8     | 1.0         | 412    | 2 A91158 | probable transport |
| 30         | 8     | 1.0         | 412    | 2 G86003 | probable transport |
| 31         | 8     | 1.0         | 412    | 2 AB1000 | type II secretion  |
| 32         | 8     | 1.0         | 469    | 2 G87513 | beta-glucosidase [ |
| 33         | 8     | 1.0         | 476    | 2 B44997 | merozoite surface  |
| 34         | 8     | 1.0         | 478    | 2 A32555 | major merozoite su |
| 35         | 8     | 1.0         | 482    | 2 A44997 | merozoite surface  |
| 36         | 8     | 1.0         | 497    | 2 AB0700 | probable outer mem |
| 37         | 8     | 1.0         | 591    | 2 S64727 | protein secretion  |
| 38         | 8     | 1.0         | 617    | 2 C95906 | hypothetical prote |
| 39         | 8     | 1.0         | 637    | 2 D82816 | fimbrial assembly  |
| 40         | 8     | 1.0         | 707    | 1 DNMS   | nucleolin - mouse  |
| 41         | 8     | 1.0         | 712    | 2 JH0148 | nucleolin - rat    |
| 42         | 8     | 1.0         | 737    | 2 I39547 | S-protein secretio |
| 43         | 8     | 1.0         | 771    | 2 S35681 | ESG protein - mous |
| 44         | 8     | 1.0         | 776    | 2 C83411 | secretion protein  |
| 45         | 8     | 1.0         | 980    | 2 AH1844 | hypothetical prote |
| 46         | 8     | 1.0         | 1335   | 2 G90975 | probable factor li |
| 47         | 8     | 1.0         | 1398   | 2 T20434 | hypothetical prote |
| 48         | 8     | 1.0         | 1718   | 2 T11638 | hypothetical prote |
| 49         | 8     | 1.0         | 1785   | 2 A45546 | major merozoite su |
| 50         | 8     | 1.0         | 2204   | 2 A70524 | probable PPS prote |
| 51         | 8     | 1.0         | 2383   | 2 D4962  | probable membrane  |
| 52         | 8     | 1.0         | 2660   | 2 E85822 | probable membrane  |
| 53         | 8     | 1.0         | 24     | 2 S07699 | T-cell receptor al |
| 54         | 7     | 0.9         | 55     | 2 A81917 | hypothetical prote |
| 55         | 7     | 0.9         | 56     | 2 A01092 | ribulose-bisphosph |
| 56         | 7     | 0.9         | 64     | 2 A81857 | hypothetical DNA-b |
| 57         | 7     | 0.9         | 78     | 2 E83446 | hypothetical prote |
| 58         | 7     | 0.9         | 83     | 2 C47188 | MHC class II histo |
| 59         | 7     | 0.9         | 84     | 2 S77876 | phosphate transpor |
| 60         | 7     | 0.9         | 104    | 2 C72482 | hypothetical prote |
| 61         | 7     | 0.9         | 105    | 2 H81004 | hypothetical prote |
| 62         | 7     | 0.9         | 109    | 2 T27861 | hypothetical prote |
| 63         | 7     | 0.9         | 111    | 2 B95857 | hypothetical prote |
| 64         | 7     | 0.9         | 129    | 2 T08084 | dynein light chain |
| 65         | 7     | 0.9         | 129    | 2 AI3357 | hypothetical prote |
| 66         | 7     | 0.9         | 130    | 2 C85701 | hypothetical prote |
| 67         | 7     | 0.9         | 130    | 2 F90843 | hypothetical prote |
| 68         | 7     | 0.9         | 130    | 2 I83571 | probable membrane  |
| 69         | 7     | 0.9         | 137    | 1 F64961 | hypothetical prote |
| 70         | 7     | 0.9         | 137    | 2 C85815 | hypothetical prote |
| 71         | 7     | 0.9         | 137    | 2 D90967 | hypothetical prote |
| 72         | 7     | 0.9         | 137    | 2 T52618 | hypothetical prote |
| 73         | 7     | 0.9         | 138    | 2 G84036 | hypothetical prote |
| 74         | 7     | 0.9         | 138    | 2 E85515 | unknown protein en |
| 75         | 7     | 0.9         | 138    | 2 B90665 | hypothetical prote |
| 76         | 7     | 0.9         | 140    | 2 D72575 | hypothetical prote |
| 77         | 7     | 0.9         | 140    | 2 S41785 | probable sugar tra |
| 78         | 7     | 0.9         | 152    | 2 C39384 | finger protein Hrf |
| 79         | 7     | 0.9         | 153    | 2 T49895 | oleosin-like prote |
| 80         | 7     | 0.9         | 156    | 2 AF2149 | hypothetical prote |
| 81         | 7     | 0.9         | 159    | 2 H69580 | acetyl-CoA carboxy |
| 82         | 7     | 0.9         | 167    | 2 A55438 | transcription fact |
| 83         | 7     | 0.9         | 170    | 2 S18064 | peroxidase (EC 1.1 |
| 84         | 7     | 0.9         | 171    | 2 S62588 | b zip transcriptio |
| 85         | 7     | 0.9         | 171    | 2 A99190 | hypothetical prote |
| 86         | 7     | 0.9         | 171    | 2 AH3096 | conserved hypotet  |
| 87         | 7     | 0.9         | 172    | 2 S75440 | adenine phosphorib |
| 88         | 7     | 0.9         | 172    | 2 T07099 | late embryogenesis |
| 89         | 7     | 0.9         | 175    | 2 AC0986 | hypothetical prote |
| 90         | 7     | 0.9         | 177    | 2 AC0954 | ATP synthase delta |
| 91         | 7     | 0.9         | 185    | 1 RKXMS1 | ribulose-bisphosph |
| 92         | 7     | 0.9         | 185    | 1 RKXMS2 | ribulose-bisphosph |
| 93         | 7     | 0.9         | 185    | 2 A32137 | histone H1-delta - |
| 94         | 7     | 0.9         | 189    | 2 H83764 | diaminobutyric aci |
| 95         | 7     | 0.9         | 194    | 2 F86455 | probable 60S ribos |
| 96         | 7     | 0.9         | 195    | 2 S71255 | ribosomal protein  |
| 97         | 7     | 0.9         | 196    | 2 A39384 | finger protein HPF |
| 98         | 7     | 0.9         | 197    | 2 G90055 | conserved hypotet  |
| 99         | 7     | 0.9         | 201    | 2 AH0288 | glutathione transf |
| 100        | 7     | 0.9         | 202    | 1 D69321 | conserved hypotet  |
| 101        | 7     | 0.9         | 205    | 2 S07283 | DNA-binding protei |
| 102        | 7     | 0.9         | 212    | 2 A28470 | histone H1 - mouse |



|     |   |     |     |   |        |                    |     |   |     |     |   |        |                    |
|-----|---|-----|-----|---|--------|--------------------|-----|---|-----|-----|---|--------|--------------------|
| 103 | 7 | 0.9 | 212 | 2 | E87486 | conserved hypothet | 176 | 7 | 0.9 | 374 | 2 | T48042 | hypothetical prote |
| 104 | 7 | 0.9 | 212 | 2 | T26782 | hypothetical prote | 177 | 7 | 0.9 | 375 | 2 | T01468 | hypothetical prote |
| 105 | 7 | 0.9 | 215 | 2 | S46966 | microfilarial shea | 178 | 7 | 0.9 | 375 | 2 | T08134 | oleosin-like prote |
| 106 | 7 | 0.9 | 216 | 2 | D70374 | phosphoribosylglyc | 179 | 7 | 0.9 | 377 | 2 | S31933 | actin - common tob |
| 107 | 7 | 0.9 | 217 | 2 | G86886 | deoxyguanosine kin | 180 | 7 | 0.9 | 377 | 2 | S20034 | actin 58 - potato  |
| 108 | 7 | 0.9 | 217 | 2 | JH0159 | histone H1d - rat  | 181 | 7 | 0.9 | 377 | 2 | S20096 | actin 75 - potato  |
| 109 | 7 | 0.9 | 219 | 1 | HSU01B | histone H1-4 [vali | 182 | 7 | 0.9 | 377 | 2 | S68108 | actin 4 - Arabidop |
| 110 | 7 | 0.9 | 219 | 2 | I49742 | histone H1 - mouse | 183 | 7 | 0.9 | 379 | 1 | F64633 | site-specific DNA- |
| 111 | 7 | 0.9 | 221 | 2 | S49482 | histone H1 - mouse | 184 | 7 | 0.9 | 380 | 2 | T01706 | hypothetical prote |
| 112 | 7 | 0.9 | 223 | 2 | A81285 | probable triose-ph | 185 | 7 | 0.9 | 381 | 2 | A71882 | type II DNA modifi |
| 113 | 7 | 0.9 | 223 | 2 | S16268 | auxin-induced prot | 186 | 7 | 0.9 | 381 | 2 | AG3148 | hypothetical prote |
| 114 | 7 | 0.9 | 223 | 2 | S16267 | auxin-induced prot | 187 | 7 | 0.9 | 382 | 2 | S08595 | TrfA transcription |
| 115 | 7 | 0.9 | 226 | 2 | A91033 | probable DNA trans | 188 | 7 | 0.9 | 387 | 2 | T24198 | hypothetical prote |
| 116 | 7 | 0.9 | 226 | 2 | B85877 | probable prophage  | 189 | 7 | 0.9 | 391 | 2 | A41795 | somatostatin recep |
| 117 | 7 | 0.9 | 228 | 2 | E87612 | cytochrome c, memb | 190 | 7 | 0.9 | 391 | 2 | C41795 | somatostatin recep |
| 118 | 7 | 0.9 | 228 | 2 | S46965 | microfilarial shea | 191 | 7 | 0.9 | 391 | 2 | A39297 | somatostatin recep |
| 119 | 7 | 0.9 | 230 | 2 | T41660 | Glutathione S-tran | 192 | 7 | 0.9 | 392 | 2 | E71633 | acetyl-CoA acetyl  |
| 120 | 7 | 0.9 | 233 | 2 | S11292 | SA85-1.1 protein - | 193 | 7 | 0.9 | 393 | 2 | F84068 | aspartate aminotra |
| 121 | 7 | 0.9 | 237 | 2 | S46964 | microfilarial shea | 194 | 7 | 0.9 | 393 | 2 | T21826 | hypothetical prote |
| 122 | 7 | 0.9 | 238 | 2 | JC4503 | macrophage maturat | 195 | 7 | 0.9 | 396 | 1 | JH0633 | cellular tumor ant |
| 123 | 7 | 0.9 | 238 | 2 | D75477 | hypothetical prote | 196 | 7 | 0.9 | 398 | 2 | AB2691 | conserved hypothet |
| 124 | 7 | 0.9 | 243 | 2 | H86487 | hypothetical prote | 197 | 7 | 0.9 | 399 | 2 | D95279 | probable alcohol l |
| 125 | 7 | 0.9 | 245 | 1 | CLH00A | complement subcomp | 198 | 7 | 0.9 | 405 | 2 | T18950 | hypothetical prote |
| 126 | 7 | 0.9 | 245 | 2 | G70090 | hypothetical prote | 199 | 7 | 0.9 | 405 | 2 | D98139 | rfe protein (AE006 |
| 127 | 7 | 0.9 | 250 | 2 | G86714 | conserved hypothet | 200 | 7 | 0.9 | 409 | 2 | H83448 | dihydrolipoamide s |
| 128 | 7 | 0.9 | 258 | 2 | G64567 | chromosome partiti | 201 | 7 | 0.9 | 410 | 2 | AB3546 | aminobutylaldehyde |
| 129 | 7 | 0.9 | 259 | 2 | G64568 | NH(3)-dependent NA | 202 | 7 | 0.9 | 411 | 2 | T25223 | hypothetical prote |
| 130 | 7 | 0.9 | 268 | 2 | B87676 | hypothetical prote | 203 | 7 | 0.9 | 412 | 2 | G87522 | hypothetical prote |
| 131 | 7 | 0.9 | 270 | 2 | E87649 | hypothetical prote | 204 | 7 | 0.9 | 413 | 2 | T52617 | hypothetical prote |
| 132 | 7 | 0.9 | 272 | 2 | B86523 | probable ABC trans | 205 | 7 | 0.9 | 416 | 2 | F89779 | hypothetical prote |
| 133 | 7 | 0.9 | 273 | 2 | A70696 | hypothetical prote | 206 | 7 | 0.9 | 420 | 2 | H75395 | ABC transporter, p |
| 134 | 7 | 0.9 | 273 | 2 | D71436 | hypothetical prote | 207 | 7 | 0.9 | 422 | 2 | S55484 | probable alcohol d |
| 135 | 7 | 0.9 | 276 | 2 | AF1277 | hypothetical prote | 208 | 7 | 0.9 | 430 | 2 | S58169 | porphobilinogen sy |
| 136 | 7 | 0.9 | 276 | 2 | AF1640 | hypothetical prote | 209 | 7 | 0.9 | 432 | 2 | A64602 | processing protein |
| 137 | 7 | 0.9 | 278 | 2 | E82862 | conjugal transfer  | 210 | 7 | 0.9 | 438 | 2 | F82944 | GTP-binding protei |
| 138 | 7 | 0.9 | 280 | 2 | A84117 | flagellar hook-bas | 211 | 7 | 0.9 | 440 | 2 | D81415 | chromosomal replic |
| 139 | 7 | 0.9 | 281 | 2 | B85018 | hypothetical prote | 212 | 7 | 0.9 | 442 | 1 | S11712 | transcription init |
| 140 | 7 | 0.9 | 284 | 2 | T34537 | hypothetical prote | 213 | 7 | 0.9 | 445 | 2 | D81716 | hypothetical prote |
| 141 | 7 | 0.9 | 285 | 2 | F72221 | conserved hypothet | 214 | 7 | 0.9 | 445 | 2 | E97320 | sugar/proton sympo |
| 142 | 7 | 0.9 | 306 | 2 | T07684 | NADH2 dehydrogenas | 215 | 7 | 0.9 | 451 | 2 | H96958 | hypothetical prote |
| 143 | 7 | 0.9 | 306 | 2 | T22650 | hypothetical prote | 216 | 7 | 0.9 | 452 | 2 | A69297 | acetyl-CoA decarbo |
| 144 | 7 | 0.9 | 307 | 2 | C82387 | transcription regu | 217 | 7 | 0.9 | 457 | 2 | H70820 | hypothetical glyci |
| 145 | 7 | 0.9 | 307 | 2 | AF2827 | RF1C protein limpo | 218 | 7 | 0.9 | 465 | 2 | A44498 | radial spoke-prote |
| 146 | 7 | 0.9 | 307 | 2 | D97605 | proteinase chain h | 219 | 7 | 0.9 | 469 | 2 | T36496 | probable dihydroli |
| 147 | 7 | 0.9 | 309 | 2 | E87671 | cytochrome c oxida | 220 | 7 | 0.9 | 472 | 2 | AD3284 | hypothetical prote |
| 148 | 7 | 0.9 | 310 | 2 | C83819 | probable acyl tran | 221 | 7 | 0.9 | 474 | 2 | E70828 | probable transcrip |
| 149 | 7 | 0.9 | 318 | 2 | B81886 | probable transposa | 222 | 7 | 0.9 | 477 | 2 | E86249 | hypothetical prote |
| 150 | 7 | 0.9 | 323 | 2 | T46671 | probable aromatase | 223 | 7 | 0.9 | 478 | 1 | S39590 | formate-dependent  |
| 151 | 7 | 0.9 | 329 | 2 | B86838 | hypothetical prote | 224 | 7 | 0.9 | 478 | 1 | DPHUM2 | methionyl aminopep |
| 152 | 7 | 0.9 | 333 | 2 | H82309 | thiosulfate ABC tr | 225 | 7 | 0.9 | 478 | 2 | D91260 | periplasmic cytoch |
| 153 | 7 | 0.9 | 333 | 2 | G36788 | hypothetical prote | 226 | 7 | 0.9 | 478 | 2 | H86100 | hypothetical prote |
| 154 | 7 | 0.9 | 337 | 2 | AB2300 | hypothetical prote | 227 | 7 | 0.9 | 478 | 2 | T35759 | pyruvate kinase -  |
| 155 | 7 | 0.9 | 340 | 2 | D70615 | antigen fbpC2 - My | 228 | 7 | 0.9 | 479 | 2 | T40683 | cell cycle protein |
| 156 | 7 | 0.9 | 341 | 2 | T04554 | probable peroxidas | 229 | 7 | 0.9 | 481 | 2 | F97472 | hypothetical prote |
| 157 | 7 | 0.9 | 342 | 2 | S63366 | probable membrane  | 230 | 7 | 0.9 | 484 | 2 | A86503 | pyruvate kinase li |
| 158 | 7 | 0.9 | 342 | 2 | S50400 | hypothetical prote | 231 | 7 | 0.9 | 484 | 2 | A72119 | pyruvate kinase Cp |
| 159 | 7 | 0.9 | 344 | 2 | T48827 | hypothetical prote | 232 | 7 | 0.9 | 484 | 2 | T06063 | hypothetical prote |
| 160 | 7 | 0.9 | 345 | 2 | S75235 | hypothetical prote | 233 | 7 | 0.9 | 485 | 2 | G71527 | probable pyruvate  |
| 161 | 7 | 0.9 | 345 | 2 | S12788 | transcription fact | 234 | 7 | 0.9 | 488 | 2 | A71337 | probable phosphogl |
| 162 | 7 | 0.9 | 345 | 2 | T15599 | hypothetical prote | 235 | 7 | 0.9 | 488 | 2 | T06259 | NADH2 dehydrogenas |
| 163 | 7 | 0.9 | 350 | 2 | E87327 | hypothetical prote | 236 | 7 | 0.9 | 488 | 2 | A42491 | NADH2 dehydrogenas |
| 164 | 7 | 0.9 | 351 | 2 | S78042 | IG mu chain C regi | 237 | 7 | 0.9 | 488 | 2 | T06260 | NADH2 dehydrogenas |
| 165 | 7 | 0.9 | 352 | 2 | C84199 | hypothetical prote | 238 | 7 | 0.9 | 489 | 2 | S25943 | NADH2 dehydrogenas |
| 166 | 7 | 0.9 | 352 | 2 | A82744 | hypothetical prote | 239 | 7 | 0.9 | 499 | 2 | S71136 | NADH2 dehydrogenas |
| 167 | 7 | 0.9 | 352 | 2 | B87507 | conserved hypothet | 240 | 7 | 0.9 | 507 | 2 | T02898 | hypothetical prote |
| 168 | 7 | 0.9 | 353 | 2 | A98131 | mocA protein limpo | 241 | 7 | 0.9 | 507 | 2 | S56143 | cell cycle protein |
| 169 | 7 | 0.9 | 353 | 2 | AH3156 | oxidoreductase moc | 242 | 7 | 0.9 | 510 | 2 | T31630 | hypothetical prote |
| 170 | 7 | 0.9 | 354 | 2 | H96928 | ABC transporter, p | 243 | 7 | 0.9 | 510 | 2 | F82566 | GuaI protein XF236 |
| 171 | 7 | 0.9 | 362 | 2 | T38464 | probable peptidyl- | 244 | 7 | 0.9 | 511 | 2 | T35194 | transcription init |
| 172 | 7 | 0.9 | 365 | 1 | W2W118 | E2 protein - human | 245 | 7 | 0.9 | 515 | 1 | DNE102 | NADH2 dehydrogenas |
| 173 | 7 | 0.9 | 368 | 2 | S36564 | E2 protein - human | 246 | 7 | 0.9 | 522 | 2 | B55481 | CAMP-dependent pro |
| 174 | 7 | 0.9 | 368 | 2 | AB3001 | histidinol-phospha | 247 | 7 | 0.9 | 523 | 2 | AC0032 | probable type III  |
| 175 | 7 | 0.9 | 368 | 2 | F98282 | histidinol-phospha | 248 | 7 | 0.9 | 525 | 1 | A75570 | 2-oxo acid dehydro |

|     |   |     |     |   |        |                     |     |   |     |       |      |         |
|-----|---|-----|-----|---|--------|---------------------|-----|---|-----|-------|------|---------|
| 249 | 7 | 0.9 | 526 | 2 | S26869 | pyruvate kinase (E  | 322 | 7 | 0.9 | 850   | 1    | T05180  |
| 250 | 7 | 0.9 | 534 | 2 | S41735 | cholesterol estera  | 323 | 7 | 0.9 | 853   | 1    | S20595  |
| 251 | 7 | 0.9 | 535 | 2 | AE3202 | hypothetical prote  | 324 | 7 | 0.9 | 870   | 2    | G86450  |
| 252 | 7 | 0.9 | 538 | 2 | S05684 | lipase I precursor  | 325 | 7 | 0.9 | 878   | 2    | T37978  |
| 253 | 7 | 0.9 | 539 | 2 | S16989 | dihydrolipoamide S  | 326 | 7 | 0.9 | 882   | 2    | A70507  |
| 254 | 7 | 0.9 | 539 | 2 | A99481 | hypothetical prote  | 327 | 7 | 0.9 | 904   | 1    | VGBBK2  |
| 255 | 7 | 0.9 | 540 | 2 | A75250 | carboxylesterase,   | 328 | 7 | 0.9 | 905   | 2    | T23229  |
| 256 | 7 | 0.9 | 544 | 2 | A81560 | Pyrophosphate-fruc  | 329 | 7 | 0.9 | 919   | 1    | PXZRP   |
| 257 | 7 | 0.9 | 544 | 2 | H86510 | fructose-6-P phosph | 330 | 7 | 0.9 | 920   | 1    | PXNCP   |
| 258 | 7 | 0.9 | 549 | 2 | JN0553 | triacylglycerol li  | 331 | 7 | 0.9 | 925   | 2    | JC2033  |
| 259 | 7 | 0.9 | 549 | 2 | JN0551 | triacylglycerol li  | 332 | 7 | 0.9 | 925   | 2    | H84582  |
| 260 | 7 | 0.9 | 549 | 2 | JN0551 | triacylglycerol li  | 333 | 7 | 0.9 | 925   | 2    | H84582  |
| 261 | 7 | 0.9 | 550 | 2 | C75557 | hypothetical prote  | 334 | 7 | 0.9 | 998   | 2    | H75005  |
| 262 | 7 | 0.9 | 551 | 2 | S09144 | ND5 intron 3 prote  | 335 | 7 | 0.9 | 1007  | 2    | T13693  |
| 263 | 7 | 0.9 | 555 | 2 | F72111 | fructose-6-P phosph | 336 | 7 | 0.9 | 1010  | 1    | PKZRP2P |
| 264 | 7 | 0.9 | 558 | 2 | T23849 | hypothetical prote  | 337 | 7 | 0.9 | 1029  | 2    | H86179  |
| 265 | 7 | 0.9 | 561 | 2 | T14792 | hypothetical prote  | 338 | 7 | 0.9 | 1072  | 2    | G95851  |
| 266 | 7 | 0.9 | 568 | 2 | A45804 | Ig mu chain C reg   | 339 | 7 | 0.9 | 1089  | 2    | T30843  |
| 267 | 7 | 0.9 | 571 | 2 | B86150 | hypothetical prote  | 340 | 7 | 0.9 | 1127  | 2    | G71274  |
| 268 | 7 | 0.9 | 572 | 2 | S49985 | dihydropyrimidinas  | 341 | 7 | 0.9 | 1132  | 2    | C75359  |
| 269 | 7 | 0.9 | 572 | 2 | JC5317 | dihydropyrimidinas  | 342 | 7 | 0.9 | 1147  | 1    | NMAX1B  |
| 270 | 7 | 0.9 | 572 | 2 | S58889 | collapsin response  | 343 | 7 | 0.9 | 1164  | 2    | S46769  |
| 271 | 7 | 0.9 | 575 | 2 | S59740 | NRD1 protein - Yea  | 344 | 7 | 0.9 | 1199  | 2    | T15826  |
| 272 | 7 | 0.9 | 578 | 2 | D82053 | fimbrial assembly   | 345 | 7 | 0.9 | 1214  | 2    | G97419  |
| 273 | 7 | 0.9 | 581 | 2 | T22341 | hypothetical prote  | 346 | 7 | 0.9 | 1219  | 2    | T14578  |
| 274 | 7 | 0.9 | 589 | 2 | T19216 | hypothetical prote  | 347 | 7 | 0.9 | 1248  | 2    | E96827  |
| 275 | 7 | 0.9 | 590 | 1 | T35297 | probable dihydroli  | 348 | 7 | 0.9 | 1248  | 2    | AH2637  |
| 276 | 7 | 0.9 | 595 | 2 | G02075 | transcription repr  | 349 | 7 | 0.9 | 1256  | 2    | C71436  |
| 277 | 7 | 0.9 | 610 | 2 | S58885 | derailed receptor   | 350 | 7 | 0.9 | 1476  | 2    | A45773  |
| 278 | 7 | 0.9 | 610 | 2 | T22687 | hypothetical prote  | 351 | 7 | 0.9 | 1476  | 2    | T01055  |
| 279 | 7 | 0.9 | 611 | 2 | T19217 | hypothetical prote  | 352 | 7 | 0.9 | 1516  | 2    | A43855  |
| 280 | 7 | 0.9 | 614 | 2 | T11390 | cell division prot  | 353 | 7 | 0.9 | 1536  | 2    | A43855  |
| 281 | 7 | 0.9 | 627 | 2 | S76378 | cell division prot  | 354 | 7 | 0.9 | 1541  | 2    | T02831  |
| 282 | 7 | 0.9 | 628 | 2 | S73248 | hypothetical prote  | 355 | 7 | 0.9 | 1567  | 2    | S11872  |
| 283 | 7 | 0.9 | 628 | 2 | AC2261 | cell division prot  | 356 | 7 | 0.9 | 1573  | 2    | T50113  |
| 284 | 7 | 0.9 | 629 | 2 | AH0521 | dihydrolipoamide S  | 357 | 7 | 0.9 | 1585  | 2    | B69948  |
| 285 | 7 | 0.9 | 630 | 1 | XXECDP | dihydrolipoamide S  | 358 | 7 | 0.9 | 1827  | 2    | T34288  |
| 286 | 7 | 0.9 | 630 | 2 | G85494 | hypothetical prote  | 359 | 7 | 0.9 | 1861  | 2    | T13845  |
| 287 | 7 | 0.9 | 630 | 2 | G90643 | hypothetical prote  | 360 | 7 | 0.9 | 1939  | 2    | T18372  |
| 288 | 7 | 0.9 | 630 | 2 | T27991 | hypothetical prote  | 361 | 7 | 0.9 | 1966  | 2    | T08991  |
| 289 | 7 | 0.9 | 634 | 2 | T17911 | hypothetical prote  | 362 | 7 | 0.9 | 2082  | 2    | T37056  |
| 290 | 7 | 0.9 | 634 | 2 | T18702 | hypothetical prote  | 363 | 7 | 0.9 | 2340  | 2    | B71754  |
| 291 | 7 | 0.9 | 641 | 2 | D87289 | hypothetical prote  | 364 | 7 | 0.9 | 2561  | 1    | L40456  |
| 292 | 7 | 0.9 | 643 | 2 | T43052 | hypothetical prote  | 365 | 7 | 0.9 | 2591  | 2    | T30288  |
| 293 | 7 | 0.9 | 646 | 2 | JN0473 | fatty acid transpo  | 366 | 7 | 0.9 | 2715  | 2    | T13049  |
| 294 | 7 | 0.9 | 655 | 2 | A54872 | p-selectin precurs  | 367 | 7 | 0.9 | 2793  | 2    | B90784  |
| 295 | 7 | 0.9 | 656 | 2 | AH2402 | acyl-CoA dehydroge  | 368 | 7 | 0.9 | 2806  | 2    | D85644  |
| 296 | 7 | 0.9 | 659 | 2 | S36551 | cell division prot  | 369 | 7 | 0.9 | 2812  | 2    | T43271  |
| 297 | 7 | 0.9 | 668 | 2 | S39836 | E1 protein - human  | 370 | 7 | 0.9 | 3131  | 2    | S39842  |
| 298 | 7 | 0.9 | 672 | 1 | A29836 | beta-galactosidase  | 371 | 7 | 0.9 | 3131  | 2    | T30132  |
| 299 | 7 | 0.9 | 677 | 2 | T11321 | NADH2 dehydrogenase | 372 | 6 | 0.8 | 10797 | 2    | T30132  |
| 300 | 7 | 0.9 | 681 | 2 | T23454 | hypothetical prote  | 373 | 6 | 0.8 | 15    | 2    | PA0084  |
| 301 | 7 | 0.9 | 687 | 2 | T49960 | hypothetical prote  | 374 | 6 | 0.8 | 15    | 2    | PA0087  |
| 302 | 7 | 0.9 | 690 | 2 | T23775 | FtsH-like protein   | 375 | 6 | 0.8 | 22    | 2    | D47256  |
| 303 | 7 | 0.9 | 698 | 2 | T23469 | hypothetical prote  | 376 | 6 | 0.8 | 24    | 2    | I61491  |
| 304 | 7 | 0.9 | 699 | 2 | H87275 | hypothetical prote  | 377 | 6 | 0.8 | 29    | 2    | S01614  |
| 305 | 7 | 0.9 | 711 | 2 | C84767 | thio-disulfide int  | 378 | 6 | 0.8 | 30    | 2    | A72205  |
| 306 | 7 | 0.9 | 719 | 2 | T27977 | lin-15A protein -   | 379 | 6 | 0.8 | 37    | 2    | B81851  |
| 307 | 7 | 0.9 | 723 | 2 | H82035 | fatty oxidation co  | 380 | 6 | 0.8 | 38    | 2    | T06938  |
| 308 | 7 | 0.9 | 725 | 2 | A90255 | hypothetical prote  | 381 | 6 | 0.8 | 40    | 2    | H71330  |
| 309 | 7 | 0.9 | 725 | 2 | T33498 | hypothetical prote  | 382 | 6 | 0.8 | 41    | 2    | E83876  |
| 310 | 7 | 0.9 | 739 | 2 | B86478 | protein F1504.15 [  | 383 | 6 | 0.8 | 42    | 2    | T16667  |
| 311 | 7 | 0.9 | 771 | 2 | G71409 | probable replicati  | 384 | 6 | 0.8 | 43    | 2    | T29051  |
| 312 | 7 | 0.9 | 782 | 2 | S50719 | hypothetical prote  | 385 | 6 | 0.8 | 45    | 2    | T20937  |
| 313 | 7 | 0.9 | 782 | 2 | JC7284 | phospholipase A2 (  | 386 | 6 | 0.8 | 48    | 2    | T13337  |
| 314 | 7 | 0.9 | 796 | 2 | T36308 | probable integral   | 387 | 6 | 0.8 | 52    | 2    | T07269  |
| 315 | 7 | 0.9 | 802 | 2 | C90326 | hypothetical prote  | 388 | 6 | 0.8 | 53    | 2    | A82773  |
| 316 | 7 | 0.9 | 813 | 2 | G86232 | protein F1086.32 [  | 389 | 6 | 0.8 | 60    | 2    | JN0741  |
| 317 | 7 | 0.9 | 815 | 2 | A80037 | hypothetical prote  | 390 | 6 | 0.8 | 60    | 2    | D92125  |
| 318 | 7 | 0.9 | 820 | 2 | G86246 | hypothetical prote  | 391 | 6 | 0.8 | 65    | 2    | S19568  |
| 319 | 7 | 0.9 | 833 | 2 | T43446 | hypothetical prote  | 392 | 6 | 0.8 | 67    | 2    | T30718  |
| 320 | 7 | 0.9 | 837 | 1 | S54423 | cuter membrane ush  | 393 | 6 | 0.8 | 68    | 2    | B84730  |
| 321 | 7 | 0.9 | 849 | 2 | S77217 | phosphorylase (EC   | 394 | 6 | 0.8 | 70    | 1    | SVFZ    |
|     |   |     |     |   |        |                     |     |   |     | 71    | SVXC |         |

|     |     |   |         |     |   |     |     |   |        |                    |
|-----|-----|---|---------|-----|---|-----|-----|---|--------|--------------------|
| 395 | 71  | 2 | AF2087  | 468 | 6 | 0.8 | 107 | 2 | JQ0143 | hypothetical 12.4K |
| 396 | 71  | 2 | AG2087  | 469 | 6 | 0.8 | 107 | 2 | T16585 | hypothetical prote |
| 397 | 72  | 2 | T11975  | 470 | 6 | 0.8 | 108 | 1 | CCNC   | cytochrome c [vali |
| 398 | 72  | 2 | F97436  | 471 | 6 | 0.8 | 108 | 2 | C89816 | hypothetical prote |
| 399 | 76  | 2 | A11993  | 472 | 6 | 0.8 | 108 | 2 | C72457 | hypothetical prote |
| 400 | 77  | 2 | A21966  | 473 | 6 | 0.8 | 109 | 1 | CCDBK  | cytochrome c [vali |
| 401 | 77  | 2 | AH1347  | 474 | 6 | 0.8 | 109 | 1 | R6TEIT | acidic ribosomal p |
| 402 | 78  | 2 | S53402  | 475 | 6 | 0.8 | 109 | 2 | E84202 | ferredoxin [import |
| 403 | 78  | 2 | F84461  | 476 | 6 | 0.8 | 110 | 1 | S40402 | protein-export pro |
| 404 | 79  | 2 | B30924  | 477 | 6 | 0.8 | 110 | 2 | S11172 | ubiquinol-cytochro |
| 405 | 79  | 2 | AB2323  | 478 | 6 | 0.8 | 110 | 2 | S58811 | cytochrome c - yea |
| 406 | 80  | 2 | S78295  | 479 | 6 | 0.8 | 110 | 2 | T01565 | protein-export pro |
| 407 | 80  | 2 | B69996  | 480 | 6 | 0.8 | 110 | 2 | F11135 | protein-export mem |
| 408 | 82  | 2 | E90324  | 481 | 6 | 0.8 | 110 | 2 | A85981 | kdgf protein - Erw |
| 409 | 82  | 1 | FDFFIAW | 482 | 6 | 0.8 | 110 | 2 | S17709 | hypothetical prote |
| 410 | 82  | 2 | S02336  | 483 | 6 | 0.8 | 110 | 2 | T17913 | hypothetical prote |
| 411 | 82  | 2 | SO30706 | 484 | 6 | 0.8 | 111 | 1 | CHL    | cytochrome c - imp |
| 412 | 82  | 2 | A05151  | 485 | 6 | 0.8 | 111 | 2 | S36447 | homeotic protein.  |
| 413 | 82  | 2 | SO7075  | 486 | 6 | 0.8 | 111 | 2 | AB3543 | hypothetical prote |
| 414 | 82  | 2 | I51125  | 487 | 6 | 0.8 | 111 | 2 | S74466 | photosystem 11 l3k |
| 415 | 82  | 2 | F90808  | 488 | 6 | 0.8 | 112 | 2 | S74466 | probable head-tail |
| 416 | 84  | 2 | B85668  | 489 | 6 | 0.8 | 112 | 2 | A98288 | high mobility grou |
| 417 | 84  | 2 | C97004  | 490 | 6 | 0.8 | 112 | 2 | A44382 | hypothetical prote |
| 418 | 84  | 2 | A64849  | 491 | 6 | 0.8 | 113 | 2 | P90319 | hypothetical prote |
| 419 | 85  | 2 | S25837  | 492 | 6 | 0.8 | 113 | 2 | JC7800 | neutrophil activat |
| 420 | 85  | 2 | D83288  | 493 | 6 | 0.8 | 113 | 2 | S07898 | endothelin 3 - rab |
| 421 | 85  | 2 | T06989  | 494 | 6 | 0.8 | 115 | 2 | F70803 | hypothetical prote |
| 422 | 85  | 2 | AB0638  | 495 | 6 | 0.8 | 115 | 2 | T41084 | probable ribosomal |
| 423 | 85  | 2 | T06988  | 496 | 6 | 0.8 | 116 | 2 | AB2553 | hypothetical prote |
| 424 | 88  | 2 | E91226  | 497 | 6 | 0.8 | 116 | 2 | AG2478 | hypothetical prote |
| 425 | 89  | 2 | E84409  | 498 | 6 | 0.8 | 116 | 2 | C95280 | hypothetical prote |
| 426 | 89  | 2 | C86073  | 499 | 6 | 0.8 | 117 | 2 | T10943 | superoxide dismuta |
| 427 | 89  | 2 | S40804  | 500 | 6 | 0.8 | 118 | 2 | G84322 | 50S ribosomal prot |
| 428 | 89  | 2 | AD0951  | 501 | 6 | 0.8 | 118 | 2 | D75141 | transcription repr |
| 429 | 89  | 2 | W3WL13  | 502 | 6 | 0.8 | 118 | 2 | I50709 | Msx-2 - chicken (f |
| 430 | 91  | 2 | T24120  | 503 | 6 | 0.8 | 118 | 2 | B64630 | hypothetical prote |
| 431 | 91  | 2 | F61996  | 504 | 6 | 0.8 | 119 | 1 | R57E14 | ribosomal protein  |
| 432 | 91  | 2 | F69252  | 505 | 6 | 0.8 | 119 | 1 | Q0B223 | BLRF3 protein - hu |
| 433 | 92  | 2 | E86723  | 506 | 6 | 0.8 | 119 | 2 | T32898 | hypothetical prote |
| 434 | 92  | 2 | S31025  | 507 | 6 | 0.8 | 120 | 2 | H90341 | partial transposas |
| 435 | 92  | 2 | S76007  | 508 | 6 | 0.8 | 120 | 2 | B95051 | conserved hypotet  |
| 436 | 92  | 2 | AB3336  | 509 | 6 | 0.8 | 121 | 2 | G97921 | alkaline shock pro |
| 437 | 92  | 2 | AD0682  | 510 | 6 | 0.8 | 121 | 2 | S48420 | probable membrane  |
| 438 | 93  | 2 | AD0682  | 511 | 6 | 0.8 | 121 | 2 | AF0323 | probable acid shoc |
| 439 | 94  | 2 | S72920  | 512 | 6 | 0.8 | 121 | 2 | A86886 | 50S ribosomal prot |
| 440 | 94  | 2 | A70540  | 513 | 6 | 0.8 | 122 | 2 | E70393 | hypothetical prote |
| 441 | 95  | 2 | A13378  | 514 | 6 | 0.8 | 122 | 2 | B71336 | hypothetical prote |
| 442 | 96  | 2 | E70913  | 515 | 6 | 0.8 | 123 | 2 | A64648 | hypothetical prote |
| 443 | 96  | 2 | S51930  | 516 | 6 | 0.8 | 123 | 2 | F95047 | probable ribosome  |
| 444 | 96  | 2 | T20661  | 517 | 6 | 0.8 | 123 | 2 | E97918 | hypothetical prote |
| 445 | 97  | 2 | S02376  | 518 | 6 | 0.8 | 123 | 2 | E83707 | homeotic protein C |
| 446 | 97  | 2 | D82789  | 519 | 6 | 0.8 | 123 | 2 | AB2467 | hypothetical prote |
| 447 | 97  | 2 | B97662  | 520 | 6 | 0.8 | 124 | 2 | D75606 | hypothetical prote |
| 448 | 98  | 2 | T11308  | 521 | 6 | 0.8 | 124 | 2 | C75565 | hypothetical prote |
| 449 | 98  | 2 | S63662  | 522 | 6 | 0.8 | 125 | 2 | G72893 | hypothetical prote |
| 450 | 100 | 2 | FC1130  | 523 | 6 | 0.8 | 125 | 2 | T43414 | DNA-directed RNA p |
| 451 | 100 | 2 | S73886  | 524 | 6 | 0.8 | 125 | 2 | AG2064 | hypothetical prote |
| 452 | 100 | 2 | H71132  | 525 | 6 | 0.8 | 125 | 2 | T04662 | high mobility grou |
| 453 | 100 | 2 | T17689  | 526 | 6 | 0.8 | 126 | 2 | G71369 | probable ribosome- |
| 454 | 101 | 2 | D82931  | 527 | 6 | 0.8 | 126 | 2 | F72729 | hypothetical prote |
| 455 | 101 | 2 | AE0539  | 528 | 6 | 0.8 | 126 | 2 | H70387 | hypothetical prote |
| 456 | 101 | 2 | AB2655  | 529 | 6 | 0.8 | 126 | 2 | C36868 | capC homolog - Xan |
| 457 | 102 | 2 | S65294  | 530 | 6 | 0.8 | 127 | 2 | H82725 | hypothetical prote |
| 458 | 102 | 2 | S65294  | 531 | 6 | 0.8 | 127 | 2 | G84316 | hypothetical prote |
| 459 | 102 | 2 | S65294  | 532 | 6 | 0.8 | 128 | 2 | A84343 | hypothetical prote |
| 460 | 103 | 2 | A97484  | 533 | 6 | 0.8 | 128 | 2 | S53584 | probable membrane  |
| 461 | 103 | 2 | A12701  | 534 | 6 | 0.8 | 129 | 2 | D49094 | methylnalonyl-CoA  |
| 462 | 103 | 2 | AF3455  | 535 | 6 | 0.8 | 129 | 2 | T47357 | hypothetical prote |
| 463 | 104 | 2 | S23528  | 536 | 6 | 0.8 | 129 | 2 | A82227 | conserved hypotet  |
| 464 | 104 | 2 | H84393  | 537 | 6 | 0.8 | 130 | 2 | T48196 | transforming growt |
| 465 | 105 | 1 | HSB011  | 538 | 6 | 0.8 | 130 | 2 | AH0888 | probable exported  |
| 466 | 106 | 1 | R6D0P2  | 539 | 6 | 0.8 | 130 | 2 | H72399 | hypothetical prote |
| 467 | 106 | 2 | T09196  | 540 | 6 | 0.8 | 132 | 2 | E98205 | hypothetical prote |
|     |     |   |         |     | 6 | 0.8 | 132 | 2 | I39773 | hypothetical prote |

|     |   |     |     |   |        |                    |     |     |     |   |        |                    |
|-----|---|-----|-----|---|--------|--------------------|-----|-----|-----|---|--------|--------------------|
| 541 | 6 | 0.8 | 133 | 2 | 578365 | H+-transporting tw | 614 | 0.8 | 148 | 2 | T21488 | hypothetical prote |
| 542 | 6 | 0.8 | 133 | 2 | F90156 | 1su ribosomal prot | 615 | 0.8 | 149 | 1 | W6WL58 | E6 protein - human |
| 543 | 6 | 0.8 | 134 | 2 | T11385 | NADH2 dehydrogenas | 616 | 0.8 | 149 | 2 | D90226 | hypothetical prote |
| 544 | 6 | 0.8 | 134 | 2 | T32065 | hypothetical prote | 617 | 0.8 | 149 | 2 | C95917 | hypothetical prote |
| 545 | 6 | 0.8 | 134 | 2 | S42040 | hypothetical prote | 618 | 0.8 | 150 | 2 | G64218 | ribosomal protein  |
| 546 | 6 | 0.8 | 134 | 2 | S09875 | hypothetical prote | 619 | 0.8 | 150 | 2 | H87250 | molybdopterin conv |
| 547 | 6 | 0.8 | 135 | 2 | G72523 | hypothetical prote | 620 | 0.8 | 151 | 2 | S62825 | ribosomal protein  |
| 548 | 6 | 0.8 | 136 | 2 | F84809 | histone H2A [impor | 621 | 0.8 | 151 | 2 | D87275 | hypothetical prote |
| 549 | 6 | 0.8 | 136 | 2 | G64815 | ybhQ protein - Esc | 622 | 0.8 | 151 | 2 | T10768 | latex allergen Hav |
| 550 | 6 | 0.8 | 136 | 2 | E90737 | hypothetical prote | 623 | 0.8 | 152 | 2 | AG2499 | probable periplasm |
| 551 | 6 | 0.8 | 136 | 2 | F85597 | hypothetical prote | 624 | 0.8 | 152 | 2 | C81189 | probable hth trans |
| 552 | 6 | 0.8 | 136 | 2 | AC0539 | probable membrane  | 625 | 0.8 | 152 | 2 | D71569 | hypothetical prote |
| 553 | 6 | 0.8 | 137 | 2 | C70481 | hypothetical prote | 626 | 0.8 | 154 | 2 | S77527 | hypothetical prote |
| 554 | 6 | 0.8 | 137 | 2 | S57695 | hypothetical prote | 627 | 0.8 | 154 | 2 | T24047 | hypothetical prote |
| 555 | 6 | 0.8 | 137 | 2 | B96603 | transcription fact | 628 | 0.8 | 154 | 2 | T47830 | hypothetical prote |
| 556 | 6 | 0.8 | 137 | 2 | C75600 | DNA-binding protei | 629 | 0.8 | 154 | 2 | AF3354 | translation initia |
| 557 | 6 | 0.8 | 137 | 2 | S00696 | vitellogenin precu | 630 | 0.8 | 154 | 2 | B97868 | hypothetical prote |
| 558 | 6 | 0.8 | 137 | 2 | A95297 | hypothetical prote | 631 | 0.8 | 154 | 2 | S64076 | exbD protein (limp |
| 559 | 6 | 0.8 | 138 | 1 | N1BPA7 | internal vitron pr | 632 | 0.8 | 155 | 2 | A99404 | T44783             |
| 560 | 6 | 0.8 | 138 | 2 | B30125 | 40S ribosomal prot | 633 | 0.8 | 155 | 2 | AC0310 | conserved hypothet |
| 561 | 6 | 0.8 | 138 | 2 | H87734 | hypothetical prote | 634 | 0.8 | 155 | 2 | S75888 | hypothetical prote |
| 562 | 6 | 0.8 | 139 | 1 | W6WL52 | E6 protein - bovin | 635 | 0.8 | 155 | 2 | B84175 | hypothetical prote |
| 563 | 6 | 0.8 | 139 | 2 | F87686 | hypothetical prote | 636 | 0.8 | 155 | 2 | S31951 | penicillin-binding |
| 564 | 6 | 0.8 | 139 | 2 | T27115 | hypothetical prote | 637 | 0.8 | 155 | 2 | JC2444 | ribosomal protein  |
| 565 | 6 | 0.8 | 140 | 2 | A86149 | actin-depolymerizi | 638 | 0.8 | 157 | 1 | T10730 | ribosomal protein  |
| 566 | 6 | 0.8 | 140 | 2 | A72667 | hypothetical prote | 639 | 0.8 | 157 | 2 | JN0549 | transcription elon |
| 567 | 6 | 0.8 | 140 | 2 | S57063 | probable membrane  | 640 | 0.8 | 157 | 2 | H87601 | hypothetical prote |
| 568 | 6 | 0.8 | 140 | 2 | A70552 | hypothetical prote | 641 | 0.8 | 157 | 2 | B64008 | hypothetical prote |
| 569 | 6 | 0.8 | 140 | 2 | T39344 | very hypothetical  | 642 | 0.8 | 157 | 2 | C81072 | hypothetical prote |
| 570 | 6 | 0.8 | 141 | 2 | A90545 | acp synthase epsil | 643 | 0.8 | 157 | 2 | D97850 | hypothetical prote |
| 571 | 6 | 0.8 | 141 | 2 | B32998 | chorion protein S1 | 644 | 0.8 | 157 | 2 | F83446 | hypothetical prote |
| 572 | 6 | 0.8 | 142 | 2 | H69870 | hypothetical prote | 645 | 0.8 | 157 | 2 | F95313 | hypothetical prote |
| 573 | 6 | 0.8 | 142 | 2 | T00616 | probable nucleic a | 646 | 0.8 | 157 | 2 | T22645 | hypothetical prote |
| 574 | 6 | 0.8 | 142 | 2 | T10078 | hypothetical prote | 647 | 0.8 | 158 | 2 | T22645 | hypothetical prote |
| 575 | 6 | 0.8 | 142 | 2 | H84856 | hypothetical prote | 648 | 0.8 | 159 | 2 | S65784 | ribosomal protein  |
| 576 | 6 | 0.8 | 142 | 2 | T43151 | transcription coac | 649 | 0.8 | 159 | 2 | H72613 | hypothetical prote |
| 577 | 6 | 0.8 | 142 | 2 | D72340 | hypothetical prote | 650 | 0.8 | 160 | 2 | F97607 | hypothetical prote |
| 578 | 6 | 0.8 | 143 | 1 | B69007 | probable transcrip | 651 | 0.8 | 160 | 2 | F95334 | Nex18 Symbioticall |
| 579 | 6 | 0.8 | 143 | 2 | E86805 | hypothetical prote | 652 | 0.8 | 160 | 2 | AG2829 | conserved hypothet |
| 580 | 6 | 0.8 | 143 | 2 | S77097 | hypothetical prote | 653 | 0.8 | 161 | 2 | S25206 | prsa protein - Esc |
| 581 | 6 | 0.8 | 143 | 2 | G72000 | Hth transcription  | 654 | 0.8 | 161 | 2 | C82747 | conserved hypothet |
| 582 | 6 | 0.8 | 143 | 2 | B96624 | HTH transcription  | 655 | 0.8 | 161 | 2 | G72711 | hypothetical prote |
| 583 | 6 | 0.8 | 143 | 2 | B96668 | high-mobility-grou | 656 | 0.8 | 161 | 2 | A69181 | hypothetical prote |
| 584 | 6 | 0.8 | 144 | 2 | A22571 | hypothetical prote | 657 | 0.8 | 161 | 2 | AG2664 | ATP synthase B cha |
| 585 | 6 | 0.8 | 144 | 2 | F87278 | HPr kinase/phospha | 658 | 0.8 | 161 | 2 | F97446 | hypothetical prote |
| 586 | 6 | 0.8 | 144 | 2 | F70707 | hypothetical prote | 659 | 0.8 | 161 | 2 | AB2054 | hypothetical prote |
| 587 | 6 | 0.8 | 145 | 2 | A84587 | hypothetical prote | 660 | 0.8 | 162 | 2 | I40185 | hypothetical prote |
| 588 | 6 | 0.8 | 145 | 2 | T33961 | hypothetical prote | 661 | 0.8 | 163 | 2 | A81786 | conserved hypothet |
| 589 | 6 | 0.8 | 145 | 2 | F83834 | hypothetical prote | 662 | 0.8 | 163 | 2 | A81210 | conserved hypothet |
| 590 | 6 | 0.8 | 145 | 2 | AF3353 | hypothetical prote | 663 | 0.8 | 164 | 2 | S40382 | hypothetical prote |
| 591 | 6 | 0.8 | 145 | 2 | F90185 | conserved hypothet | 664 | 0.8 | 164 | 2 | B64628 | transcription elon |
| 592 | 6 | 0.8 | 145 | 2 | B89114 | protein 2K742.5 [i | 665 | 0.8 | 164 | 2 | D70130 | conserved hypothet |
| 593 | 6 | 0.8 | 145 | 2 | S77736 | probable zinc fing | 666 | 0.8 | 165 | 2 | S04044 | embryonic abundan  |
| 594 | 6 | 0.8 | 146 | 1 | A48897 | aminoglycoside N6' | 667 | 0.8 | 165 | 2 | F97716 | hypothetical prote |
| 595 | 6 | 0.8 | 146 | 2 | AD2257 | transposase alr361 | 668 | 0.8 | 166 | 2 | T33056 | hypothetical prote |
| 596 | 6 | 0.8 | 146 | 2 | AE2142 | transposase alr269 | 669 | 0.8 | 166 | 2 | G75519 | conserved hypothet |
| 597 | 6 | 0.8 | 146 | 2 | AF2450 | transposase alr515 | 670 | 0.8 | 167 | 1 | JC1102 | endothelin 3 precu |
| 598 | 6 | 0.8 | 146 | 2 | AG1808 | transposase all001 | 671 | 0.8 | 167 | 2 | T29649 | hypothetical prote |
| 599 | 6 | 0.8 | 146 | 2 | AG2355 | transposase all439 | 672 | 0.8 | 167 | 2 | B72112 | hypothetical prote |
| 600 | 6 | 0.8 | 146 | 2 | AG2360 | transposase all443 | 673 | 0.8 | 167 | 2 | D86511 | hypothetical prote |
| 601 | 6 | 0.8 | 146 | 2 | AH2407 | transposase all481 | 674 | 0.8 | 167 | 2 | AF3206 | acetyltransferase  |
| 602 | 6 | 0.8 | 146 | 2 | A13340 | hypothetical prote | 675 | 0.8 | 168 | 2 | S33613 | ribulose-bisphosph |
| 603 | 6 | 0.8 | 147 | 2 | E82251 | hypothetical prote | 676 | 0.8 | 168 | 2 | I40352 | single-stranded DN |
| 604 | 6 | 0.8 | 147 | 2 | G64057 | biopolymer transpo | 677 | 0.8 | 168 | 2 | AB3362 | single-strand bind |
| 605 | 6 | 0.8 | 147 | 2 | C84231 | hypothetical prote | 678 | 0.8 | 168 | 2 | H83109 | hypothetical prote |
| 606 | 6 | 0.8 | 147 | 2 | B70436 | hypothetical prote | 679 | 0.8 | 168 | 2 | D75565 | conserved hypothet |
| 607 | 6 | 0.8 | 147 | 2 | F75013 | hypothetical prote | 680 | 0.8 | 168 | 2 | A45943 | vitelline membrane |
| 608 | 6 | 0.8 | 148 | 2 | E64939 | hypothetical prote | 681 | 0.8 | 169 | 2 | T34224 | hypothetical prote |
| 609 | 6 | 0.8 | 148 | 2 | B90941 | hypothetical prote | 682 | 0.8 | 170 | 2 | B72122 | ribosomal protein  |
| 610 | 6 | 0.8 | 148 | 2 | F85789 | hypothetical prote | 683 | 0.8 | 170 | 2 | G86500 | 110 ribosomal prot |
| 611 | 6 | 0.8 | 148 | 2 | AH0712 | hypothetical prote | 684 | 0.8 | 170 | 2 | G87153 | 50S ribosomal prot |
| 612 | 6 | 0.8 | 148 | 2 | G87262 | probable membrane  | 685 | 0.8 | 170 | 2 | F87482 | hypothetical prote |
| 613 | 6 | 0.8 | 148 | 2 | E75283 | transcription regu | 686 | 0.8 | 170 | 2 | B72252 | cob(I)alamin adeno |

|     |     |   |        |                     |     |   |     |     |   |        |                     |
|-----|-----|---|--------|---------------------|-----|---|-----|-----|---|--------|---------------------|
| 687 | 170 | 2 | A13334 | D-ribose-binding p  | 760 | 6 | 0.8 | 190 | 2 | E95420 | hypothetical prote  |
| 688 | 171 | 2 | F69366 | conserved hypotet   | 761 | 6 | 0.8 | 190 | 2 | A85360 | hypothetical prote  |
| 689 | 171 | 2 | B87628 | conserved hypotet   | 762 | 6 | 0.8 | 191 | 2 | D90712 | lipote biosynthes   |
| 690 | 171 | 2 | F87298 | hypothetical prote  | 763 | 6 | 0.8 | 191 | 2 | H85562 | lipote biosynthes   |
| 691 | 171 | 2 | S35569 | sex-determining pr  | 764 | 6 | 0.8 | 191 | 2 | D64797 | lipote-protein li   |
| 692 | 171 | 2 | T31478 | hypothetical prote  | 765 | 6 | 0.8 | 191 | 2 | T13536 | hypothetical prote  |
| 693 | 171 | 2 | G95316 | hypothetical prote  | 766 | 6 | 0.8 | 192 | 2 | E82542 | two-component syst  |
| 694 | 172 | 1 | Q93ED7 | HSLF4 protein - hu  | 767 | 6 | 0.8 | 192 | 2 | G82162 | conserved hypotet   |
| 695 | 172 | 2 | A24019 | nonhistone chromos  | 768 | 6 | 0.8 | 193 | 1 | JN0728 | hypothetical prote  |
| 696 | 172 | 2 | G81015 | probable gluconoki  | 769 | 6 | 0.8 | 193 | 2 | S19978 | ribosomal protein   |
| 697 | 173 | 2 | JV5554 | hypothetical prote  | 770 | 6 | 0.8 | 193 | 2 | A83152 | conserved hypotet   |
| 698 | 173 | 2 | JV0047 | hypothetical prote  | 771 | 6 | 0.8 | 193 | 2 | H86404 | probable lipid tra  |
| 699 | 173 | 2 | B90812 | probable membrane   | 772 | 6 | 0.8 | 193 | 2 | S32992 | hypothetical prote  |
| 700 | 173 | 2 | F85671 | probable membrane   | 773 | 6 | 0.8 | 193 | 2 | F81779 | probable transposa  |
| 701 | 173 | 2 | AG0641 | conserved hypotet   | 774 | 6 | 0.8 | 194 | 1 | HSTR1  | histone H1 - trout  |
| 702 | 174 | 2 | B69977 | NAD(P)H oxidoreduc  | 775 | 6 | 0.8 | 194 | 2 | I50710 | fibroblast growth   |
| 703 | 174 | 2 | A70447 | conserved hypotet   | 776 | 6 | 0.8 | 194 | 2 | S07930 | hypothetical prote  |
| 704 | 174 | 2 | B38880 | pulB protein - Kle  | 777 | 6 | 0.8 | 194 | 2 | T30660 | hypothetical prote  |
| 705 | 174 | 2 | H62659 | hypothetical prote  | 778 | 6 | 0.8 | 195 | 2 | T14965 | hypothetical prote  |
| 706 | 174 | 2 | A96030 | hypothetical prote  | 779 | 6 | 0.8 | 195 | 2 | E98022 | hypothetical prote  |
| 707 | 175 | 2 | A71227 | probable cob(II)ala | 780 | 6 | 0.8 | 195 | 2 | C95156 | conserved hypotet   |
| 708 | 175 | 2 | G75194 | hypothetical prote  | 781 | 6 | 0.8 | 195 | 2 | AH3213 | conserved hypotet   |
| 709 | 175 | 2 | A97857 | hypothetical prote  | 782 | 6 | 0.8 | 196 | 1 | A65105 | phosphoheptose iso  |
| 710 | 176 | 1 | KLSWM  | calcium-binding pr  | 783 | 6 | 0.8 | 196 | 2 | F91132 | hypothetical prote  |
| 711 | 176 | 2 | T41361 | hypothetical prote  | 784 | 6 | 0.8 | 196 | 2 | A85978 | hypothetical prote  |
| 712 | 176 | 2 | T4216  | probable V-ATPase,  | 785 | 6 | 0.8 | 196 | 2 | A08899 | probable phospho    |
| 713 | 177 | 2 | E72160 | thymidine kinase (  | 786 | 6 | 0.8 | 196 | 2 | T08137 | chlorophyll a/b-b1  |
| 714 | 177 | 2 | G97097 | diverged enzyme re  | 787 | 6 | 0.8 | 196 | 2 | JN0721 | ATP-corrinoid aden  |
| 715 | 178 | 2 | T23601 | hypothetical prote  | 788 | 6 | 0.8 | 196 | 2 | A64875 | cob(II)alamin adeno |
| 716 | 178 | 2 | B83378 | hypothetical prote  | 789 | 6 | 0.8 | 196 | 2 | AH0269 | cob(II)alamin adeno |
| 717 | 178 | 2 | G70134 | flagellar protein   | 790 | 6 | 0.8 | 196 | 2 | B99859 | cob(II)alamin adeno |
| 718 | 178 | 2 | A86370 | hypothetical prote  | 791 | 6 | 0.8 | 196 | 2 | E85760 | cob(II)alamin adeno |
| 719 | 179 | 2 | JC1487 | chymotrypsin inhib  | 792 | 6 | 0.8 | 196 | 2 | AB0654 | COB(II) alamin aden |
| 720 | 179 | 2 | B81685 | hypothetical prote  | 793 | 6 | 0.8 | 197 | 2 | T43358 | hypothetical prote  |
| 721 | 179 | 2 | F83813 | phosphatidylester   | 794 | 6 | 0.8 | 197 | 2 | T46344 | hypothetical prote  |
| 722 | 180 | 2 | H82356 | hypothetical prote  | 795 | 6 | 0.8 | 197 | 2 | C64301 | hypothetical prote  |
| 723 | 180 | 2 | H81023 | conserved hypotet   | 796 | 6 | 0.8 | 198 | 2 | JC1457 | ryudocan precursor  |
| 724 | 180 | 2 | S56943 | heat shock protein  | 797 | 6 | 0.8 | 198 | 2 | G69053 | conserved hypotet   |
| 725 | 181 | 2 | C81041 | pius assembly pro   | 798 | 6 | 0.8 | 198 | 2 | A97395 | hypothetical prote  |
| 726 | 182 | 2 | T14862 | reverse transcript  | 799 | 6 | 0.8 | 198 | 2 | AC2613 | conserved hypotet   |
| 727 | 182 | 2 | A70804 | probable lpgE prot  | 800 | 6 | 0.8 | 198 | 2 | JT0356 | Cop protein - Cios  |
| 728 | 182 | 2 | D82925 | hypothetical prote  | 801 | 6 | 0.8 | 199 | 2 | C75363 | hydrolase family p  |
| 729 | 182 | 2 | D84228 | hypothetical prote  | 802 | 6 | 0.8 | 199 | 2 | A75191 | hypothetical prote  |
| 730 | 182 | 2 | T41779 | FGF orf32 - Bombyx  | 803 | 6 | 0.8 | 199 | 2 | T47023 | hypothetical prote  |
| 731 | 182 | 2 | B70315 | hypothetical prote  | 804 | 6 | 0.8 | 199 | 2 | AB0236 | probable Na(+)-tra  |
| 732 | 183 | 2 | S57772 | early nodulin GRP3  | 805 | 6 | 0.8 | 199 | 2 | G72261 | conserved hypotet   |
| 733 | 183 | 2 | S57773 | early nodulin GRP3  | 806 | 6 | 0.8 | 199 | 2 | S75871 | hypothetical prote  |
| 734 | 183 | 2 | T14653 | hypothetical prote  | 807 | 6 | 0.8 | 199 | 2 | H90205 | hypothetical prote  |
| 735 | 183 | 2 | B70751 | hypothetical prote  | 808 | 6 | 0.8 | 199 | 2 | F90119 | hypothetical prote  |
| 736 | 183 | 2 | E95872 | conserved hypotet   | 809 | 6 | 0.8 | 200 | 1 | B69834 | phage-related rpl   |
| 737 | 183 | 2 | H81485 | hypothetical prote  | 810 | 6 | 0.8 | 200 | 2 | AC3509 | dephospho-CoA kina  |
| 738 | 184 | 2 | S10357 | luciferin-binding   | 811 | 6 | 0.8 | 200 | 2 | D75057 | pyroglutaryl-pepti  |
| 739 | 184 | 2 | G89177 | conserved hypotet   | 812 | 6 | 0.8 | 200 | 2 | B83642 | hypothetical prote  |
| 740 | 184 | 2 | H81850 | probable membrane   | 813 | 6 | 0.8 | 201 | 1 | T29447 | probable bacitraci  |
| 741 | 184 | 2 | F81090 | hypothetical prote  | 814 | 6 | 0.8 | 201 | 2 | T47113 | protocatechuate 3,  |
| 742 | 184 | 2 | T03178 | probable zinc fing  | 815 | 6 | 0.8 | 201 | 2 | A10409 | superoxide dismuta  |
| 743 | 184 | 2 | D72307 | conserved hypotet   | 816 | 6 | 0.8 | 201 | 2 | A91001 | hypothetical prote  |
| 744 | 185 | 2 | B81559 | hypothetical prote  | 817 | 6 | 0.8 | 201 | 2 | A93779 | hypothetical prote  |
| 745 | 185 | 2 | A75636 | conserved hypotet   | 818 | 6 | 0.8 | 201 | 2 | B85846 | unknown protein en  |
| 746 | 185 | 2 | T30309 | hypothetical prote  | 819 | 6 | 0.8 | 201 | 2 | D85640 | hypothetical prote  |
| 747 | 185 | 2 | E86625 | transcription regu  | 820 | 6 | 0.8 | 201 | 2 | AG1668 | hypothetical prote  |
| 748 | 185 | 2 | G70199 | hypothetical prote  | 821 | 6 | 0.8 | 201 | 2 | A11296 | hypothetical prote  |
| 749 | 186 | 2 | C86861 | hypothetical prote  | 822 | 6 | 0.8 | 201 | 2 | G70874 | probable transcrip  |
| 750 | 186 | 2 | T45938 | hypothetical prote  | 823 | 6 | 0.8 | 201 | 2 | G85736 | hypothetical prote  |
| 751 | 186 | 2 | T22646 | hypothetical prote  | 824 | 6 | 0.8 | 201 | 2 | B83319 | hypothetical prote  |
| 752 | 186 | 2 | C75558 | acetyl-CoA carboxy  | 825 | 6 | 0.8 | 201 | 2 | E75567 | hypothetical prote  |
| 753 | 187 | 2 | AH1290 | polypeptide deform  | 826 | 6 | 0.8 | 201 | 2 | G71428 | hypothetical prote  |
| 754 | 187 | 2 | C70609 | hypothetical prote  | 827 | 6 | 0.8 | 202 | 2 | AG2776 | ATP-dependent Clp   |
| 755 | 187 | 2 | T47342 | hypothetical prote  | 828 | 6 | 0.8 | 202 | 2 | E97556 | clpp (AF218420) [i  |
| 756 | 187 | 2 | S38246 | hypothetical prote  | 829 | 6 | 0.8 | 202 | 2 | D89843 | hypothetical prote  |
| 757 | 189 | 2 | AE1446 | probable scaffoldi  | 830 | 6 | 0.8 | 202 | 2 | JC4990 | chymotrypsin inhib  |
| 758 | 189 | 2 | T35951 | probable SIR2 fami  | 831 | 6 | 0.8 | 203 | 2 | B14716 | GTP-binding protei  |
| 759 | 190 | 2 |        |                     | 832 | 6 | 0.8 | 203 | 2 | G91066 | hypothetical prote  |

|     |     |     |   |     |     |   |        |                     |
|-----|-----|-----|---|-----|-----|---|--------|---------------------|
| 833 | 203 | 0.8 | 6 | 906 | 216 | 2 | C75403 | hypothetical prote  |
| 834 | 203 | 0.8 | 6 | 907 | 216 | 2 | AH3497 | hypothetical cytos  |
| 835 | 203 | 0.8 | 6 | 908 | 216 | 2 | D97868 | hypothetical prote  |
| 836 | 204 | 0.8 | 6 | 909 | 217 | 2 | C64359 | ribosomal protein   |
| 837 | 204 | 0.8 | 6 | 910 | 217 | 2 | A83146 | lipote-protein li   |
| 838 | 204 | 0.8 | 6 | 911 | 217 | 2 | S29309 | hypothetical prote  |
| 839 | 204 | 0.8 | 6 | 912 | 217 | 2 | A46630 | jacalin precursor   |
| 840 | 204 | 0.8 | 6 | 913 | 218 | 2 | S36974 | hypothetical prote  |
| 841 | 204 | 0.8 | 6 | 914 | 218 | 2 | H69060 | chromosome partiti  |
| 842 | 204 | 0.8 | 6 | 915 | 218 | 2 | AD0112 | ribose-5-phosphate  |
| 843 | 205 | 0.8 | 6 | 916 | 218 | 2 | C75394 | ATP phosphoribosyl  |
| 844 | 205 | 0.8 | 6 | 917 | 218 | 2 | T35174 | hypothetical prote  |
| 845 | 205 | 0.8 | 6 | 918 | 218 | 2 | AG1725 | carboxylesterase (  |
| 846 | 205 | 0.8 | 6 | 919 | 218 | 2 | TU0277 | sulfide reductase,  |
| 847 | 205 | 0.8 | 6 | 920 | 218 | 2 | A40347 | hypothetical prote  |
| 848 | 206 | 0.8 | 6 | 921 | 218 | 2 | AG1725 | hypothetical prote  |
| 849 | 206 | 0.8 | 6 | 922 | 219 | 2 | S55840 | orotate phosphorib  |
| 850 | 206 | 0.8 | 6 | 923 | 219 | 2 | T52087 | glutathione transf  |
| 851 | 206 | 0.8 | 6 | 924 | 219 | 2 | T52087 | glutathione transf  |
| 852 | 207 | 0.8 | 6 | 925 | 219 | 2 | T52088 | glutathione transf  |
| 853 | 207 | 0.8 | 6 | 926 | 219 | 2 | T52088 | glutathione transf  |
| 854 | 207 | 0.8 | 6 | 927 | 219 | 2 | T52086 | conserved hypoteth  |
| 855 | 207 | 0.8 | 6 | 928 | 219 | 2 | B90238 | ribose-5-phosphate  |
| 856 | 207 | 0.8 | 6 | 929 | 219 | 2 | A65076 | ribosephosphate is  |
| 857 | 207 | 0.8 | 6 | 930 | 219 | 2 | A98102 | ribosephosphate is  |
| 858 | 207 | 0.8 | 6 | 931 | 219 | 2 | E85947 | ribose-5-phosphate  |
| 859 | 207 | 0.8 | 6 | 932 | 219 | 2 | A80874 | hypothetical prote  |
| 860 | 207 | 0.8 | 6 | 933 | 219 | 2 | B72662 | hypothetical prote  |
| 861 | 208 | 0.8 | 6 | 934 | 219 | 2 | H64973 | hypothetical prote  |
| 862 | 208 | 0.8 | 6 | 935 | 219 | 2 | A98989 | hypothetical prote  |
| 863 | 208 | 0.8 | 6 | 936 | 219 | 2 | C85834 | hypothetical prote  |
| 864 | 209 | 0.8 | 6 | 937 | 219 | 2 | B75533 | hypothetical prote  |
| 865 | 209 | 0.8 | 6 | 938 | 219 | 2 | F03131 | conserved hypoteth  |
| 866 | 209 | 0.8 | 6 | 939 | 219 | 2 | E75477 | hypothetical prote  |
| 867 | 209 | 0.8 | 6 | 940 | 220 | 2 | A46597 | acidic calmodulin-  |
| 868 | 209 | 0.8 | 6 | 941 | 220 | 2 | T30688 | hypothetical prote  |
| 869 | 209 | 0.8 | 6 | 942 | 220 | 2 | S76623 | O-methyltransferas  |
| 870 | 210 | 0.8 | 6 | 943 | 220 | 2 | B90393 | hypothetical prote  |
| 871 | 210 | 0.8 | 6 | 944 | 220 | 2 | H70568 | hypothetical prote  |
| 872 | 210 | 0.8 | 6 | 945 | 221 | 1 | F64137 | regulatory protein  |
| 873 | 210 | 0.8 | 6 | 946 | 221 | 1 | AC0272 | probable transactin |
| 874 | 210 | 0.8 | 6 | 947 | 221 | 2 | T35855 | probable ABC trans  |
| 875 | 210 | 0.8 | 6 | 948 | 221 | 2 | C87214 | conserved hypoteth  |
| 876 | 210 | 0.8 | 6 | 949 | 221 | 2 | T50665 | hypothetical prote  |
| 877 | 211 | 0.8 | 6 | 950 | 222 | 1 | MNNZSP | nonstructural prot  |
| 878 | 211 | 0.8 | 6 | 951 | 222 | 2 | S78174 | probable transport  |
| 879 | 211 | 0.8 | 6 | 952 | 222 | 2 | G83125 | conserved hypoteth  |
| 880 | 211 | 0.8 | 6 | 953 | 222 | 2 | T27916 | hypothetical prote  |
| 881 | 212 | 0.8 | 6 | 954 | 222 | 2 | H72662 | hypothetical prote  |
| 882 | 212 | 0.8 | 6 | 955 | 222 | 2 | S68832 | hypothetical prote  |
| 883 | 212 | 0.8 | 6 | 956 | 223 | 2 | AF0161 | probable deoxyribo  |
| 884 | 213 | 0.8 | 6 | 957 | 223 | 2 | S32872 | regulatory protein  |
| 885 | 213 | 0.8 | 6 | 958 | 223 | 2 | B71057 | DNA mismatch repai  |
| 886 | 213 | 0.8 | 6 | 959 | 223 | 2 | E83596 | hypothetical prote  |
| 887 | 213 | 0.8 | 6 | 960 | 223 | 2 | S16269 | auxin-induced prot  |
| 888 | 213 | 0.8 | 6 | 961 | 223 | 2 | D83878 | O-methyltransferas  |
| 889 | 213 | 0.8 | 6 | 962 | 223 | 2 | AG3288 | 2-hydroxy-6-oxo-2   |
| 890 | 213 | 0.8 | 6 | 963 | 224 | 1 | JQ0597 | nonstructural prot  |
| 891 | 213 | 0.8 | 6 | 964 | 224 | 1 | A46343 | nonstructural prot  |
| 892 | 213 | 0.8 | 6 | 965 | 224 | 2 | F71678 | CbDiacylglycerol-   |
| 893 | 213 | 0.8 | 6 | 966 | 224 | 2 | F69091 | ribosomal protein   |
| 894 | 213 | 0.8 | 6 | 967 | 224 | 2 | B97200 | response regulator  |
| 895 | 214 | 0.8 | 6 | 968 | 224 | 2 | AC3181 | two component resp  |
| 896 | 214 | 0.8 | 6 | 969 | 224 | 2 | AI1021 | siderophore biosyn  |
| 897 | 215 | 0.8 | 6 | 970 | 224 | 2 | S20463 | diol dehydrogenase  |
| 898 | 215 | 0.8 | 6 | 971 | 224 | 2 | AD0760 | hypothetical prote  |
| 899 | 215 | 0.8 | 6 | 972 | 224 | 2 | H71203 | hypothetical prote  |
| 900 | 215 | 0.8 | 6 | 973 | 224 | 2 | G75218 | hypothetical prote  |
| 901 | 215 | 0.8 | 6 | 974 | 224 | 2 | G70119 | hypothetical prote  |
| 902 | 216 | 0.8 | 6 | 975 | 224 | 2 | T32655 | hypothetical prote  |
| 903 | 216 | 0.8 | 6 | 976 | 225 | 1 | A95394 | hypothetical prote  |
| 904 | 216 | 0.8 | 6 | 977 | 225 | 1 | R3MX5  | ribosomal protein   |
| 905 | 216 | 0.8 | 6 | 978 | 225 | 2 | S29953 | surface glycoprote  |
|     |     |     |   |     |     |   | S29950 | surface glycoprote  |

```
979      6 0.8 225 2 C85437
980      6 0.8 225 2 C84310
981      6 0.8 225 2 S01073
982      6 0.8 226 1 XJBYS5
983      6 0.8 226 2 C64356
984      6 0.8 226 2 AC0634
985      6 0.8 226 2 H82861
986      6 0.8 226 2 H69727
987      6 0.8 227 1 PHXJ7
988      6 0.8 227 2 C82152
989      6 0.8 227 2 T03807
990      6 0.8 227 2 A70885
991      6 0.8 228 2 T36379
992      6 0.8 228 2 T47294
993      6 0.8 228 2 G97002
994      6 0.8 228 2 RH0031
995      6 0.8 229 1 C71435
996      6 0.8 229 2 C87577
997      6 0.8 229 2 T15359
998      6 0.8 230 2 G98992
999      6 0.8 230 2 AB1462
1000     6 0.8 230 2 AC1099

ALIGNMENTS

RESULT 1
A:1985
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: A81985
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel,
; Holroyd, S.; Jorgensen, K.; Leather, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A8175; MUID:20222556; PMID:10761919
A:Accession: A81985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-761 <PAR>
A:Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CA883938.1; PID:g737938
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: pilQ; NWA0650

Query Match 71.3%; Score 548; DB 2; Length 761;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 222 PAKQWIDFRKQGNAGIIEAALGFAGQPDISQQHDIIVTLKNHTLPTTLQRLSDVAD 281
DB 214 PAKQWIDFRKQGNAGIIEAALGFAGQPDISQQHDIIVTLKNHTLPTTLQRLSDVAD 273

QY 282 FKTPOKVTLKRNNDTQIIITAGNWLWNKSAAPGYFTFOVLPPKKQNLSSGGVNNAPK 341
DB 274 FKTPOKVTLKRNNDTQIIITAGNWLWNKSAAPGYFTFOVLPPKKQNLSSGGVNNAPK 333

QY 342 TFGTKISLDFQDVEITLQILAKESGNNIVASDSVNGKMTLSLKDVPDQALDLVMOA 401
DB 334 TFGTKISLDFQDVEITLQILAKESGNNIVASDSVNGKMTLSLKDVPDQALDLVMOA 393

QY 402 RNLDMEQGNVNIAPRELLAKDKAFLOAKDIDALGALYONFOLKYNVEFRSLR 461
DB 394 RNLDMEQGNVNIAPRELLAKDKAFLOAKDIDALGALYONFOLKYNVEFRSLR 453

QY 462 LDNADTGNHTLVSGRGSLVLPDNTLITVTRSVIEKFKRLIDELDVPQAQVMIEAR 521
DB 454 LDNADTGNHTLVSGRGSLVLPDNTLITVTRSVIEKFKRLIDELDVPQAQVMIEAR 513

QY 522 IVEAADGFSRDLGVKFGATGKKLKNDTSAFGWGVNSGFGGDDKWAETKINLPITAAAN 581
```

```
514 IVEAADGFSRDLGVKFGATGKKLKNDTSAFGWGVNSGFGGDDKWAETKINLPITAAAN 573
582 SISLVRAISSGALNLELSASESLSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANG 641
574 SISLVRAISSGALNLELSASESLSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANG 633
642 GSSTNTELLKAVLGITVTPNITPDGIIIMTVKINKDSPAQCSAGNQTILCISTKLNLTQA 701
634 GSSTNTELLKAVLGITVTPNITPDGIIIMTVKINKDSPAQCSAGNQTILCISTKLNLTQA 693
702 MYENGSTLIVGGIYEDNGNTLTVPVLLGDIPVIGNLFTKRGKTKDRRELLIFITPRIMG 761
694 MYENGSTLIVGGIYEDNGNTLTVPVLLGDIPVIGNLFTKRGKTKDRRELLIFITPRIMG 753
762 TAGNSLRV 769
754 TAGNSLRV 761

RESULT 2
S70838
pilQ protein - Neisseria gonorrhoeae
C:Species: Neisseria gonorrhoeae
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 08-Oct-1999
C:Accession: S70838; S70835
R:Drake, S.L.; Komey, M.
submitted to the EMBL Data Library, November 1995
A:Description: The product of the pilQ gene is essential for the biogenesis of type I
A:Reference number: S70838
A:Accession: S70838
A:Molecule type: DNA
A:Residues: 1-720 <DRA>
A:Cross-references: EMBL:U40596; NID:g1173872; PIDN:AAC43603.1; PID:g1173875
R:Drake, S.L.; Komey, M.
Mol. Microbiol. 18, 975-986, 1995
A:Title: The product of the pilQ gene is essential for the biogenesis of type IV pilI
A:Reference number: S70835; MUID:96422484; PMID:8825101
A:Accession: S70835
A:Molecule type: DNA
A:Residues: 140-214 <DRW>
A:Cross-references: EMBL:U40596; NID:g1173872; PIDN:AAC43603.1; PID:g1173875
A:Experimental source: strain MS11
C:Genetics:
A:Gene: pilQ

Query Match 19.5%; Score 150; DB 2; Length 720;
Best Local Similarity 100.0%; Pred. No. 9.6e-141;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNTKLTIIISGLFVATAAFOTASAGNIITDKVSSLPNKQKIVKVSFDKEIVNFTGVTSS 60
DB 1 MNTKLTIIISGLFVATAAFOTASAGNIITDKVSSLPNKQKIVKVSFDKEIVNFTGVTSS 60

QY 61 PARIALDFEQTGISMDQVLEYADPLLSKISAAQNSRARLVLNLPKPGQYNTVEVRGNKV 120
DB 61 PARIALDFEQTGISMDQVLEYADPLLSKISAAQNSRARLVLNLPKPGQYNTVEVRGNKV 120

QY 121 WIFINESDDTVSAPARPAVKAAPAAKQ 150
DB 121 WIFINESDDTVSAPARPAVKAAPAAKQ 150

RESULT 3
A37051
outer membrane protein-molecular complex (omc) precursor - Neisseria gonorrhoeae
C:Species: Neisseria gonorrhoeae
C:Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 17-Feb-1994
C:Accession: A37051
R:Tsal, W.M.; Larsen, S.H.; Wilde III, C.E.
Infect. Immun. 57, 2653-2659, 1989
A:Title: Cloning and DNA sequence of the omc gene encoding the outer membrane protein
A:Reference number: A37051; MUID:89339707; PMID:2503445
A:Accession: A37051
```

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-711 <TSA>  
A:Cross-references: GB:U19944; GB:M22564  
A:Keywords: membrane protein

Query Match 18.5%; Score 142; DB 2; Length 711;  
Best Local Similarity 100.0%; Pred. No. 9.1e-133;  
Matches 142; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNTKLTIIISGLFVATAAFTAGAGNITDIKYSLLPNKQIKIVKVSFDKEIVNPTGVTSS 60  
DB 1 MNTKLTIIISGLFVATAAFTAGAGNITDIKYSLLPNKQIKIVKVSFDKEIVNPTGVTSS 60

QY 61 PARIALDFEQTGISMDOOVLEYADPLLSKISAQNSRRARLVNLNKPQYNTVRGNKV 120  
DB 61 PARIALDFEQTGISMDOOVLEYADPLLSKISAQNSRRARLVNLNKPQYNTVRGNKV 120

QY 121 WIFINESDDTVSAPARPAVAAA 142  
DB 121 WIFINESDDTVSAPARPAVAAA 142

RESULT 4  
B34469  
A:Title: Secretion protein pult precursor - Klebsiella pneumoniae (strain UNF5023)  
C:Species: Klebsiella pneumoniae  
C:Date: 15-Jun-1990 #sequence\_revision 31-Dec-1993 #text\_change 20-Feb-1995  
C:Accession: B34469; B31394  
R:Denfert, C.; Reyss, I.; Wandersman, C.; Pugsley, A.P.  
J. Biol. Chem. 267, 17462-17468, 1992  
A:Title: Protein secretion by gram-negative bacteria. Characterization of two membrane proteins  
A:Reference number: A34469; MUID:90008916; PMID:2677007  
A:Accession: B34469  
A:Molecule type: DNA  
A:Residues: 1-660 <DEN>  
A:Cross-references: GB:M32613  
C:Genetics:  
A:Gene: pult  
A:Start codon: TTG

Query Match 1.6%; Score 12; DB 2; Length 660;  
Best Local Similarity 100.0%; Pred. No. 0.0039;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDIPVIG 736  
DB 559 KVPLLGDIPVIG 570

RESULT 5  
H4067  
A:Title: Hypothetical protein H10435 (transformation locus) - Haemophilus influenzae (strain Rd K)  
N:Alternate names: ORF protein  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 29-Sep-1999  
C:Accession: H4067; JH0434  
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kurlavsky, A.; Gocayne, J.D.; Scott, J.; Shiley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: H4067  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-445 <TIG>  
A:Cross-references: GB:U32726; GB:L42023; NID:G1573399; PIDN:AAC22094.1; PID:G1573410; T  
A:Experimental source: strain Rd KW20  
R:Tomb, J.F.; El-Hajj, H.; Smith, H.O. Gene 104, 1-10, 1991  
A:Title: Nucleotide sequence of a cluster of genes involved in the transformation of Haemophilus influenzae

A:Reference number: JH0430; MUID:92009183; PMID:1916268  
A:Accession: JH0434  
A:Molecule type: DNA  
A:Residues: 1-432, 'NVRGVEV', 440, 'K' <TOM>  
A:Cross-references: GB:M62809; NID:G148990; PIDN:AAA25012.1; PID:G148997  
A:Experimental source: strain Rd  
C:Superfamily: hypothetical protein H10435

Query Match 1.4%; Score 11; DB 2; Length 445;  
Best Local Similarity 100.0%; Pred. No. 0.028;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDIPVI 735  
DB 393 KVPLLGDIPVI 403

RESULT 6  
S37345  
A:Title: pilQ protein - Pseudomonas aeruginosa  
C:Species: Pseudomonas aeruginosa  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 03-Feb-1994  
C:Accession: S37345  
R:Martin, P.R.; Hobbs, M.; Free, P.D.; Jeske, Y.; Mattick, J.S. Mol. Microbiol. 9, 857-868, 1993  
A:Title: Characterization of pilQ, a new gene required for the biogenesis of type 4 fimbriae  
A:Reference number: S37345; MUID:94049125; PMID:7901733  
A:Accession: S37345  
A:Molecule type: DNA  
A:Residues: 1-714 <MAR>  
A:Cross-references: GB:L13865  
C:Genetics:  
A:Gene: pilQ

Query Match 1.4%; Score 11; DB 2; Length 714;  
Best Local Similarity 100.0%; Pred. No. 0.042;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 515 QVMIEARIVEA 525  
DB 449 QVMIEARIVEA 459

RESULT 7  
A83016  
A:Title: type 4 fimbrial biogenesis protein PilQ PA5040 [imported] - Pseudomonas aeruginosa (strain PA5040)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: A83016  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Adam, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lory, S.; Olson, M.V. Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: A83016  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-714 <STO>  
A:Cross-references: GB:AE004917; GB:AE004091; NID:G9951326; PIDN:AA08425.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: pilQ; PA5040

Query Match 1.4%; Score 11; DB 2; Length 714;  
Best Local Similarity 100.0%; Pred. No. 0.042;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 515 QVMIEARIVEA 525  
DB 449 QVMIEARIVEA 459





```

probable general secretion pathway protein d precursor - Escherichia coli (strain K-12
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: H65125
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:9742617; PMID:9278503
A:Accession: H65125
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-654 <BLAT>
A:Cross-references: GB:AE000409; GB:U00096; NID:gl789718; PIDN:AACT6350.1; PID:gl78972
A:Experimental source: strain K-12, substrain M61655
C:Genetics:
A:Gene: ynfF

Query Match 1.2%; Score 9; DB 2; Length 654;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDP 733
DB 563 KVPLLGDP 571

RESULT 16
S39653
xcpQ protein - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 31-Dec-2000
C:Accession: S39653; A83259
R:Akrim, M.; Bally, M.; Ball, G.; Tommassen, J.; Teerink, H.; Filloux, A.; Lazdunski,
Mol. Microbiol. 10, 431-443, 1993
A:Title: Xcp-mediated protein secretion in Pseudomonas aeruginosa: identification of t
A:Reference number: S39652; MUID:95020542; PMID:7934833
A:Accession: S39653
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-658 <AKR>
A:Cross-references: EMBL:X68594; NID:9431183; PIDN:CAA48582.1; PID:9431185
A:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.;
Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L;
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pat
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: A83259
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-658 <STO>
A:Cross-references: GB:AE004734; GB:AE004091; NID:99949204; PIDN:AAG06493.1; GSPDB:GN
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: xcpQ; PA3105

Query Match 1.2%; Score 9; DB 2; Length 658;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDP 733
DB 569 KVPLLGDP 577

RESULT 17
F70352
conserved hypothetical protein aq_585 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C:Accession: F70352
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;

```

---

```

Db 674 DTRSVIEKFR 683
|||||
|||||

RESULT 13
C40361
virC-region hypothetical protein yscC precursor - Yersinia enterocolitica plasmid pYV
C:Species: Yersinia enterocolitica
C:Date: 21-Feb-1992 #sequence_revision 21-Feb-1992 #text_change 08-Oct-1999
C:Accession: C40361
R:Wachtele, T.; Vanooteghem, J.C.; Lambert de Rouvroit, C.; China, B.; Gustin, A.; Boudi
J. Bacteriol. 179, 4994-5009, 1991
A:Title: Analysis of virC, an operon involved in the secretion of Yop proteins by Yersin
A:Reference number: A40361; MUID:91317716; PMID:1860816
A:Accession: C40361
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-607 <MTC>
A:Cross-references: GB:M74011; NID:g155549; PIDN:AAC37020.1; PID:g155552
C:Genetics:
A:Genome: plasmid

Query Match 1.2%; Score 9; DB 2; Length 607;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDP 733
DB 474 KVPLLGDP 482

RESULT 14
T43573
type III secretion protein yscC - Yersinia pestis plasmid pCD1
N:Alternate names: translocation protein C homolog
C:Species: Yersinia pestis
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T43573; T42868
R:Hu, P.; Elliott, J.; McCreedy, P.; Skowronski, E.; Garnes, J.; Kobayashi, A.; Brubaker
J. Bacteriol. 180, 5192-5202, 1998
A:Title: Structural organization of virulence-associated plasmids of Yersinia pestis.
A:Reference number: T22573; MUID:98422474; PMID:9748454
A:Accession: T43573
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-607 <HUP>
A:Cross-references: EMBL:AF053946; NID:g2996222; PIDN:AAC62552.1; PID:g2996229
A:Experimental source: strain KIM
R:Perry, R.D.; Straley, S.C.; Fetherston, J.D.; Rose, D.J.; Gregor, J.; Blattner, F.R.
Infect. Immun. 66, 4611-4623, 1998
A:Title: DNA sequencing and analysis of the low-Ca2+-response plasmid pCD1 of Yersinia P
A:Reference number: T22273; MUID:98427122; PMID:9746557
A:Accession: T42868
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-607 <PER>
A:Cross-references: EMBL:AF074612; NID:g3822037; PIDN:AAC69781.1; PID:g3822061
A:Experimental source: strain KIMS
C:Genetics:
A:Gene: yscC
A:Genome: plasmid pCD1

Query Match 1.2%; Score 9; DB 2; Length 607;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDP 733
DB 474 KVPLLGDP 482

RESULT 15
H65125

```

V. Nature 392, 353-358, 1998  
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
 A:Reference number: A70300; MUID:98196666; PMID:9537320  
 A:Accession: F70352  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-705 <AQF>  
 A:Cross-references: GB:AE000697; NID:g2983212; PID:AA06820.1; PID:g2983222; GB:AE000695  
 A:Experimental source: strain VFS  
 C:Genetics:  
 A:Gene: aq\_585

Query Match 1.2%; Score 9; DB 2; Length 705;  
 Best Local Similarity 100.0%; Pred. No. 4.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 749 RELIIFITP 757  
 |||||  
 DB 691 RELIIFITP 699

RESULT 18  
 S28014  
 outD protein - Erwinia chrysanthemi  
 C:Species: Erwinia chrysanthemi  
 C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 21-Jul-2000  
 C:Accession: S28014; S23886  
 R:Condemine, G.; Dorel, C.; Hugouvieux-Cotte-Pattat, N.; Robert-Baudouy, J.  
 Mol. Microbiol. 6, 3199-3211, 1992  
 A:Title: Some of the out genes involved in the secretion of pectate lyases in Erwinia ch  
 A:Reference number: S28011; MUID:93086427; PMID:1453958  
 A:Accession: S28014  
 A:Molecule type: DNA  
 A:Residues: 1-710 <CON>  
 A:Cross-references: ENBL:X65265; NID:g3152953; PIDN:CAA46370.1; PID:g581156  
 C:Genetics:  
 A:Gene: outD  
 A:Start codon: GTG

Query Match 1.2%; Score 9; DB 2; Length 710;  
 Best Local Similarity 100.0%; Pred. No. 4.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDIP 733  
 |||||  
 DB 612 KVPLLGDIP 620

RESULT 19  
 B47021  
 pectic enzyme secretion protein OutD - Erwinia chrysanthemi  
 C:Species: Erwinia chrysanthemi  
 C:Date: 24-Feb-1994 #sequence\_revision 18-Nov-1994 #text\_change 18-Nov-1994  
 C:Accession: B47021  
 R:Lindeberg, M.; Collmer, A.  
 J. Bacteriol. 174, 7385-7397, 1992  
 A:Title: Analysis of eight out genes in a cluster required for pectic enzyme secretion  
 A:Reference number: A47021; MUID:93054355; PMID:1429461  
 A:Accession: B47021  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-712 <LIN>  
 A:Note: sequence extracted from NCBI backbone (NCBIP:118271)

Query Match 1.2%; Score 9; DB 2; Length 712;  
 Best Local Similarity 100.0%; Pred. No. 4.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDIP 733  
 |||||  
 DB 614 KVPLLGDIP 622

RESULT 20  
 A45243  
 envelope protein HrpH - Pseudomonas syringae  
 C:Species: Pseudomonas syringae  
 C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Mar-1998  
 C:Accession: A45243  
 R:Huang, H.C.; He, S.Y.; Bauer, D.W.; Collmer, A.  
 J. Bacteriol. 174, 6878-6885, 1992  
 A:Title: The Pseudomonas syringae pv. syringae 61 hrpH product, an envelope protein  
 A:Reference number: A45243; MUID:93015750; PMID:1400238  
 A:Accession: A45243  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1-748 <HUA>  
 A:Cross-references: GB:L01064; NID:gl51265; PID:gl51266  
 A:Experimental source: pv. syringae 61  
 A:Note: sequence extracted from NCBI backbone (NCBIP:116798)

Query Match 1.2%; Score 9; DB 2; Length 748;  
 Best Local Similarity 100.0%; Pred. No. 4.3;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDIP 733  
 |||||  
 DB 502 KVPLLGDIP 510

RESULT 21  
 G70772  
 hypothetical protein Rv1291c - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C:Accession: G70772  
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
 A:Reference number: A70500; MUID:98295987; PMID:9634230  
 A:Accession: G70772  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-111 <COL>  
 A:Cross-references: GB:Z73419; GB:AL123456; NID:g3261573; PIDN:CAA97761.1; PID:g32615  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: Rv1291c

Query Match 1.0%; Score 8; DB 2; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 8.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 LAALGFPAG 249  
 |||||  
 DB 22 LAALGFPAG 29

RESULT 22  
 A81036  
 mannose-1-phosphate guanylttransferase-related protein NMB1841 [imported] - Neisseria  
 C:Species: Neisseria meningitidis  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
 C:Accession: A81036  
 R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,  
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B  
 ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizza, M.  
 Science 287, 1809-1815, 2000  
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;  
 A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
 A:Reference number: A81000; MUID:20175755; PMID:10710307  
 A:Accession: A81036

A/REBIQUES: I-401 <SCH>  
A/Cross-references: EMBL:X61652: NID:g21225: PIDN:CAA4383.1; PID:g21226

C;Superfamily: porphobilinogen synthase  
C;Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 1.0%; Score 8; DB 2; Length 401;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPAK 148  
DB 53 AAPAAPAK 60

## RESULT 28

B65134  
protein transport protein hofQ precursor - Escherichia coli (strain K-12)  
C;Species: Escherichia coli  
C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C;Accession: B65134  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of Escherichia coli K-12.  
A;Reference number: A64720; MUID:97426617; PMID:9278503  
A;Accession: B65134  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-412 <BLAT>  
A;Cross-references: GB:AE000414; GB:U00096; NID:gl789783; PIDN:AAC76416.1; PID:gl789793;  
A;Experimental source: strain K-12, substrain MG1655  
C;Genetics:  
A;Gene: hofQ  
C;Superfamily: hypothetical protein HI0435

Query Match 1.0%; Score 8; DB 2; Length 412;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 726 VPLLGDP 733  
DB 375 VPLLGDP 382

## RESULT 29

A91158  
probable transport portein ECs4233 [imported] - Escherichia coli (strain O157:H7, substrain  
C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C;Accession: A91158  
R;Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic  
A;Reference number: A99629; MUID:21156231; PMID:11258796  
A;Accession: A91158  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-412 <HAY>  
A;Cross-references: PIDN:BA037656.1; PID:gl3363707; GSPDB:GN00154  
A;Experimental source: strain O157:H7, substrain RMD 0509952  
C;Genetics:  
A;Gene: ECs4233  
C;Superfamily: hypothetical protein HI0435

Query Match 1.0%; Score 8; DB 2; Length 412;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 726 VPLLGDP 733  
DB 375 VPLLGDP 382

## RESULT 30

G86003  
probable transport portein hofQ [imported] - Escherichia coli (strain O157:H7, substrain  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C;Accession: G86003  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,  
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamocousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11266551  
A;Accession: G86003  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-412 <STO>  
A;Cross-references: GB:AE005174; NID:gl2518014; PIDN:AAG58491.1; GSPDB:GN00145; UMGW:  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: hofQ  
C;Superfamily: hypothetical protein HI0435

Query Match 1.0%; Score 8; DB 2; Length 412;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 726 VPLLGDP 733  
DB 375 VPLLGDP 382

## RESULT 31

AB1000  
type II secretion system protein [imported] - Salmonella enterica subsp. enterica ser  
C;Species: Salmonella enterica subsp. enterica serovar Typhi  
C;Note: this species has also been called Salmonella typhi  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C;Accession: AB1000  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church,  
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr,  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se  
A;Reference number: AB0502; MUID:21534947; PMID:11677608  
A;Accession: AB1000  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-412 <PAR>  
A;Cross-references: GB:AL513382; PIDN:CAD08126.1; PID:gl6505105; GSPDB:GN00176  
C;Genetics:  
A;Gene: hofQ  
C;Superfamily: hypothetical protein HI0435

Query Match 1.0%; Score 8; DB 2; Length 412;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 726 VPLLGDP 733  
DB 375 VPLLGDP 382

## RESULT 32

G87513  
beta-glucosidase [imported] - Caulobacter crescentus  
C;Species: Caulobacter crescentus  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C;Accession: G87513  
R;Niernman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; K  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A;Title: Complete Genome Sequence of Caulobacter crescentus.  
A;Reference number: A87249; MUID:21173698; PMID:11259647

A44997  
 merozoite surface antigen 1 - Plasmodium chabaudi adami (strain DK)  
 C:Species: Plasmodium chabaudi adami  
 C>Date: 14-May-1993 #sequence\_revision 14-May-1993 #text\_change 11-Jan-2000  
 C:Accession: A44997  
 R:Lew, A.M.; Beck, D.J.  
 Mol. Biochem. Parasitol. 42, 153-154, 1990  
 A>Title: The epitope of a protective monoclonal antibody occurs in a region of microhete  
 A:Reference number: A44997; MUID:91042831; PMID:1700297  
 A:Accession: A44997  
 A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual t  
 A:Molecule type: DNA  
 A:Residues: 1-482 <LEW>  
 A:Cross-references: GB:M34047  
 C:Superfamily: major merozoite surface antigen  
 C:Keywords: surface antigen

Query Match 1.0%; Score 8; DB 2; Length 482;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 VATAAFQT 21  
 |||||  
 Db 36 VATAAFQT 43

RESULT 33  
 B44997  
 merozoite surface antigen 1 - Plasmodium chabaudi chabaudi (strain CB)  
 C:Species: Plasmodium chabaudi chabaudi  
 C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 11-Jan-2000  
 C:Accession: B44997  
 R:Lew, A.M.; Beck, D.J.  
 Mol. Biochem. Parasitol. 42, 153-154, 1990  
 A>Title: The epitope of a protective monoclonal antibody occurs in a region of microhete  
 A:Reference number: A44997; MUID:91042831; PMID:1700297  
 A:Accession: B44997  
 A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
 A:Molecule type: DNA  
 A:Residues: 1-476 <LEW>  
 A:Cross-references: GB:M34255  
 C:Superfamily: major merozoite surface antigen  
 C:Keywords: surface antigen

Query Match 1.0%; Score 8; DB 2; Length 476;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPAK 148  
 |||||  
 Db 128 AAPAAPAK 135

RESULT 34  
 A32555  
 major merozoite surface antigen precursor - Plasmodium chabaudi adami (fragment)  
 C:Species: Plasmodium chabaudi adami  
 C>Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 11-Jan-2000  
 C:Accession: A32555  
 R:Lew, A.M.; Langford, C.J.; Anders, R.F.; Kemp, D.J.; Saul, A.; Fardoulis, C.; Geysen,  
 Proc. Natl. Acad. Sci. U.S.A. 86, 3769-3772, 1989  
 A>Title: A protective monoclonal antibody recognizes a linear epitope in the precursor b  
 A:Reference number: A32555; MUID:89264504; PMID:2471191  
 A:Accession: A32555  
 A:Molecule type: mRNA  
 A:Residues: 1-478 <LEW>  
 A:Cross-references: GB:J04568; NID:9160419; PID:g552206  
 C:Superfamily: major merozoite surface antigen  
 C:Keywords: Glycoprotein; surface antigen  
 F:135,299/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.0%; Score 8; DB 2; Length 478;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPAK 148  
 |||||  
 Db 102 AAPAAPAK 109

RESULT 35

A44997  
 merozoite surface antigen 1 - Plasmodium chabaudi adami (strain DK)  
 C:Species: Plasmodium chabaudi adami  
 C>Date: 14-May-1993 #sequence\_revision 14-May-1993 #text\_change 11-Jan-2000  
 C:Accession: A44997  
 R:Lew, A.M.; Beck, D.J.  
 Mol. Biochem. Parasitol. 42, 153-154, 1990  
 A>Title: The epitope of a protective monoclonal antibody occurs in a region of microhe  
 A:Reference number: A44997; MUID:91042831; PMID:1700297  
 A:Accession: A44997  
 A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual t  
 A:Molecule type: DNA  
 A:Residues: 1-482 <LEW>  
 A:Cross-references: GB:M34047  
 C:Superfamily: major merozoite surface antigen  
 C:Keywords: surface antigen

Query Match 1.0%; Score 8; DB 2; Length 482;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPAK 148  
 |||||  
 Db 121 AAPAAPAK 128

RESULT 36  
 AB0700  
 probable outer membrane secretory protein spIA [imported] - Salmonella enterica subsp  
 C:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A>Note: this species has also been called Salmonella typhi  
 C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C:Accession: AB0700  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church  
 et al., T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farra  
 S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K  
 A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica ser  
 A:Reference number: AB0502; MUID:21534947; PMID:11677608  
 A:Accession: AB0700  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-497 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD01971.1; PID:g16502813; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: spIA

Query Match 1.0%; Score 8; DB 2; Length 497;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 727 PLLGDIPV 734  
 |||||  
 Db 457 PLLGDIPV 464

RESULT 37  
 S64727  
 protein secretion protein xcpQ precursor - Pseudomonas putida  
 C:Species: Pseudomonas putida  
 C>Date: 06-Dec-1996 #sequence\_revision 06-Dec-1996 #text\_change 21-Jul-2000  
 C:Accession: S64727; S47504  
 R:de Groot, A.; Krijger, J.J.; Filloux, A.; Tommassen, J.  
 Mol. Gen. Genet. 250, 491-504, 1996  
 A>Title: Characterization of type II protein secretion (xcp) genes in the plant growth  
 A:Reference number: S64724; MUID:96186881; PMID:8602167  
 A:Accession: S64727  
 A:Molecule type: DNA  
 A:Residues: 1-591 <DEG>  
 A:Cross-references: EMBL:X81085; NID:g3293032; PIDN:CAA56979.1; PID:g531740  
 C:Genetics:  
 A:Gene: xcpQ

F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-591/Product: protein secretion protein xcpQ #status predicted <MAT>  
Query Match 1.0%; Score 8; DB 2; Length 591;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 726 VPLUGDIP 733  
DB 504 VPLUGDIP 511  
RESULT 38  
C95906  
hypothetical protein [imported] - Sinorhizobium meliloti (strain 1021) megaplasmid pSymB  
C:Species: Sinorhizobium meliloti  
C>Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C:Accession: C95906  
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo  
A:Reference number: A95842; MUID:21396508; PMID:11481431  
A:Accession: C95906  
A>Status: preliminary  
A:Molecule type: DNA  
A:Cross-references: GB:AL591985; PTDN:CA48915.1; PID:gl5140388; GSPDB:GN00167  
A:Experimental source: strain 1021, megaplasmid pSymB  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Leilaure,  
heubalt, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A:Reference number: A96039; MUID:21369234; PMID:11474104  
A:Contents: annotation  
C:Genetics:  
A:Gene: SM20536  
A:Genome: plasmid  
Query Match 1.0%; Score 8; DB 2; Length 617;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 241 ELAALGFA 248  
DB 114 ELAALGFA 121  
RESULT 39  
D82816  
fimbrial assembly protein XF0373 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: D82816  
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717; PMID:10910347  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: D82816  
A>Status: preliminary  
A:Molecule type: DNA  
A:Cross-references: GB:AE003888; GB:AE003849; NID:g9105187; PTDN:AAF83183.1; GSPDB:GN001  
R:Simpon, A.J.G.; Reinach, F.C.; Aruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H  
de-Neto, E.; Docena, C.; El-Porcy, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to Genbank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins,  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, F.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawat  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silve  
M.; Teuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF0373  
Query Match 1.0%; Score 8; DB 2; Length 637;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 389 VPWDQALD 396  
DB 240 VPWDQALD 247  
RESULT 40  
DNMS  
nucleolin - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 22-Jun-1999  
C:Accession: A29958; A40769; A56240; I84688  
R:Bourbon, H.M.; Lapeyre, B.; Anallric, F.  
J. Mol. Biol. 200, 627-636, 1988  
A:Title: Structure of the mouse nucleolin gene. The complete sequence reveals that ea  
A:Reference number: A29958; MUID:88316930; PMID:3137346  
A:Accession: A29958  
A:Molecule type: DNA  
A:Residues: 1-707 <BOU>  
A:Cross-references: GB:X07699; NID:953453; PTDN:CAA30538.1; PID:G53454  
R:Pasernack, M.S.; Bleier, K.J.; McInerney, T.N.  
J. Biol. Chem. 266, 14703-14708, 1991  
A:Title: Granzyme A binding to target cell proteins. Granzyme A binds to and cleaves  
A:Reference number: A40769; MUID:91317840; PMID:1860869  
A:Accession: A40769  
A:Molecule type: protein  
A:Residues: 2-20, X' 22-24 <PAS>  
R:Yang, T.H.; Tsai, W.H.; Lee, Y.M.; Lei, H.Y.; Lai, M.Y.; Chen, D.S.; Yeh, N.H.; Lee  
Mol. Cell. Biol. 14, 6068-6074, 1994  
A:Title: Purification and characterization of nucleolin and its identification as a t  
A:Reference number: A56240; MUID:94344117; PMID:8065340  
A:Accession: A56240  
A:Molecule type: protein  
A:Residues: 2-19, 558-567 <YAN>  
R:Bourbon, H.  
Gene 68, 73-84, 1988  
A:Title: Sequence and structure of the nucleolin promoter in rodents: Characterizatio  
A:Reference number: I49118; MUID:89121496; PMID:2906027  
A:Accession: I84688  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-44 <RES>  
A:Cross-references: GB:M22089; NID:g200111; PTDN:AAA39841.1; PID:G554246  
C:Comment: This housekeeping protein is involved in the synthesis, packaging, and mat  
C:Genetics:  
A:Introns: 6/3; 45/3; 211/1; 273/1; 302/1; 349/2; 391/1; 432/3; 484/1; 525/2; 566/1;  
A:Superfamily: nucleolin; ribonucleoprotein repeat homology  
C:Keywords: DNA binding; duplication; nucleus; phosphoprotein; RNA binding; transcript  
P:310-375/Domain: ribonucleoprotein repeat homology <RRM1>  
P:310-316/Region: RNA-binding RNP2 motif  
P:349-356/Region: RNA-binding RNP1 motif  
P:396-458/Domain: ribonucleoprotein repeat  
P:397-402/Region: RNA-binding RNP2 motif  
P:431-438/Region: RNA-binding RNP1 motif  
P:488-551/Domain: ribonucleoprotein repeat homology <RRM3>  
P:489-494/Region: RNA-binding RNP2 motif  
P:524-531/Region: RNA-binding RNP1 motif  
P:570-634/Domain: ribonucleoprotein repeat homology <RRM4>  
P:571-576/Region: RNA-binding RNP2 motif

F;607-614/Region: RNA-binding RNPI motif

|                       |        |                 |               |             |
|-----------------------|--------|-----------------|---------------|-------------|
| Query Match           | 1.0%   | Score 8;        | DB 1;         | Length 707; |
| Best Local Similarity | 100.0% | Pred. No. 41;   |               |             |
| Matches               | 8;     | Conservative 0; | Mismatches 0; | Indels 0;   |
|                       |        |                 | Gaps          | 0;          |

  

|    |     |          |     |
|----|-----|----------|-----|
| QY | 140 | KAAPAAPA | 147 |
|    |     |          |     |
| Db | 181 | KAAPAAPA | 188 |

  

Search completed: December 9, 2003, 10:36:13  
Job time : 43 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

QM protein - protein search, using sw model

Run on: December 9, 2003, 10:28:28 ; Search time 17 Seconds  
(without alignments)  
2127.267 Million cell updates/sec

Title: US-09-701-271a-2

Perfect score: 769

Sequence: 1 MNTKTKIISGLFVATAAFQ.....ELLIPITPRIMTAGNSURY 769

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 142   | 18.5        | 711    | 1  | OMC_NEIGO   |
| 2          | 12    | 1.6         | 660    | 1  | GSPD_KLEPN  |
| 3          | 11    | 1.4         | 445    | 1  | COME_HAEIN  |
| 4          | 11    | 1.4         | 714    | 1  | PILQ_PSEAE  |
| 5          | 10    | 1.3         | 650    | 1  | GSPD_ERWCA  |
| 6          | 10    | 1.3         | 678    | 1  | GSPD_AERHY  |
| 7          | 10    | 1.3         | 678    | 1  | GSPD_AERSA  |
| 8          | 9     | 1.2         | 607    | 1  | YSCC_YEREN  |
| 9          | 9     | 1.2         | 650    | 1  | GSPD_ECOLI  |
| 10         | 9     | 1.2         | 658    | 1  | GSPD_PSEAE  |
| 11         | 9     | 1.2         | 674    | 1  | GSPD_VIECH  |
| 12         | 9     | 1.2         | 701    | 1  | HRPH_PSESY  |
| 13         | 9     | 1.2         | 710    | 1  | GSQD_ERWCH  |
| 14         | 9     | 1.2         | 712    | 1  | GSQD_ERWCH  |
| 15         | 8     | 1.0         | 65     | 1  | LHA3_RHOPA  |
| 16         | 8     | 1.0         | 111    | 1  | YC91_MYCTU  |
| 17         | 8     | 1.0         | 412    | 1  | HOFO_ECOLI  |
| 18         | 8     | 1.0         | 417    | 1  | HEM2_SELMA  |
| 19         | 8     | 1.0         | 706    | 1  | NUCL_MOUSE  |
| 20         | 8     | 1.0         | 712    | 1  | NUCL_RAT    |
| 21         | 8     | 1.0         | 823    | 1  | DMSA_RHOCA  |
| 22         | 8     | 1.0         | 836    | 1  | KAB7_SCHPO  |
| 23         | 8     | 1.0         | 2358   | 1  | YEEU_ECOLI  |
| 24         | 8     | 1.0         | 2660   | 1  | YEEU_ECOS7  |
| 25         | 7     | 0.9         | 129    | 1  | DYLA_CHURE  |
| 26         | 7     | 0.9         | 130    | 1  | SIB2_ECOLI  |
| 27         | 7     | 0.9         | 137    | 1  | YEDX_ECOS7  |
| 28         | 7     | 0.9         | 137    | 1  | YEDX_ECOLI  |
| 29         | 7     | 0.9         | 140    | 1  | YOR1_CALSR  |
| 30         | 7     | 0.9         | 159    | 1  | BCCP_BACSU  |
| 31         | 7     | 0.9         | 166    | 1  | THI2_HUMAN  |
| 32         | 7     | 0.9         | 167    | 1  | HESS_MOUSE  |
| 33         | 7     | 0.9         | 170    | 1  | PER2_HORVU  |

|     |   |     |     |   |                    |
|-----|---|-----|-----|---|--------------------|
| 34  | 7 | 0.9 | 171 | 1 | PCR1_SCHPO         |
| 35  | 7 | 0.9 | 172 | 1 | APT_SYN3           |
| 36  | 7 | 0.9 | 175 | 1 | YMF3_CABEL         |
| 37  | 7 | 0.9 | 185 | 1 | HID_STRPU          |
| 38  | 7 | 0.9 | 185 | 1 | RBS1_CHLRE         |
| 39  | 7 | 0.9 | 185 | 1 | RBS2_CHLRE         |
| 40  | 7 | 0.9 | 194 | 1 | RL9_ARATH          |
| 41  | 7 | 0.9 | 204 | 1 | OGG1_THEAC         |
| 42  | 7 | 0.9 | 204 | 1 | RS3A_METWA         |
| 43  | 7 | 0.9 | 205 | 1 | ERF_EPP22          |
| 44  | 7 | 0.9 | 211 | 1 | H12_MOUSE          |
| 45  | 7 | 0.9 | 211 | 1 | H12_RAT            |
| 46  | 7 | 0.9 | 218 | 1 | H14_HUMAN          |
| 47  | 7 | 0.9 | 218 | 1 | H14_MOUSE          |
| 48  | 7 | 0.9 | 220 | 1 | H13_MOUSE          |
| 49  | 7 | 0.9 | 223 | 1 | GTX1_TOBAC         |
| 50  | 7 | 0.9 | 223 | 1 | QO3662 nicotiana t |
| 51  | 7 | 0.9 | 223 | 1 | PIV6_ADEG1         |
| 52  | 7 | 0.9 | 227 | 1 | GPV_BHILO          |
| 53  | 7 | 0.9 | 230 | 1 | GTI2_SCHPO         |
| 54  | 7 | 0.9 | 236 | 1 | CYH2_BACSU         |
| 55  | 7 | 0.9 | 238 | 1 | MNDP_HUMAN         |
| 56  | 7 | 0.9 | 245 | 1 | CIQA_HUMAN         |
| 57  | 7 | 0.9 | 259 | 1 | NADE_METUA         |
| 58  | 7 | 0.9 | 271 | 1 | ECB1_CANAL         |
| 59  | 7 | 0.9 | 333 | 1 | VG24_HSV11         |
| 60  | 7 | 0.9 | 340 | 1 | A85C_MYCTU         |
| 61  | 7 | 0.9 | 342 | 1 | AR34_YEAST         |
| 62  | 7 | 0.9 | 342 | 1 | YM22_YEAST         |
| 63  | 7 | 0.9 | 345 | 1 | CEBB_HUMAN         |
| 64  | 7 | 0.9 | 348 | 1 | CEBB_BOVIN         |
| 65  | 7 | 0.9 | 348 | 1 | HIS8_CHLTE         |
| 66  | 7 | 0.9 | 352 | 1 | EGSA_HALNI         |
| 67  | 7 | 0.9 | 362 | 1 | FRH_SCHPO          |
| 68  | 7 | 0.9 | 365 | 1 | VE2_HPV18          |
| 69  | 7 | 0.9 | 368 | 1 | HIS8_AGR75         |
| 70  | 7 | 0.9 | 368 | 1 | VE2_HPV45          |
| 71  | 7 | 0.9 | 369 | 1 | CHEB_VIBPA         |
| 72  | 7 | 0.9 | 369 | 1 | T1B_PARTE          |
| 73  | 7 | 0.9 | 371 | 1 | CAR3_PVRPU         |
| 74  | 7 | 0.9 | 375 | 1 | ACT1_NAEPO         |
| 75  | 7 | 0.9 | 377 | 1 | ACT1_TOBAC         |
| 76  | 7 | 0.9 | 377 | 1 | ACT3_SOLTU         |
| 77  | 7 | 0.9 | 377 | 1 | ACT4_ARATH         |
| 78  | 7 | 0.9 | 377 | 1 | ACT7_SOLTU         |
| 79  | 7 | 0.9 | 382 | 1 | TRPA_ECOLI         |
| 80  | 7 | 0.9 | 384 | 1 | SL18_HUMAN         |
| 81  | 7 | 0.9 | 389 | 1 | VPD_BP186          |
| 82  | 7 | 0.9 | 391 | 1 | SRI_HUMAN          |
| 83  | 7 | 0.9 | 391 | 1 | SRI_MOUSE          |
| 84  | 7 | 0.9 | 391 | 1 | SRI_RAT            |
| 85  | 7 | 0.9 | 396 | 1 | P53_MESAU          |
| 86  | 7 | 0.9 | 422 | 1 | ADHF_SCHPO         |
| 87  | 7 | 0.9 | 426 | 1 | HEM1_XANAC         |
| 88  | 7 | 0.9 | 426 | 1 | HEM1_XANCH         |
| 89  | 7 | 0.9 | 430 | 1 | HEM2_PHPA          |
| 90  | 7 | 0.9 | 432 | 1 | HEM1_XANCP         |
| 91  | 7 | 0.9 | 438 | 1 | TRME_UREPA         |
| 92  | 7 | 0.9 | 440 | 1 | DNAA_CAMJE         |
| 93  | 7 | 0.9 | 451 | 1 | Y479_CLOAB         |
| 94  | 7 | 0.9 | 465 | 1 | LIPP_RAT           |
| 95  | 7 | 0.9 | 465 | 1 | RSP4_CHLRE         |
| 96  | 7 | 0.9 | 478 | 1 | AMP2_MOUSE         |
| 97  | 7 | 0.9 | 478 | 1 | AMP2_HUMAN         |
| 98  | 7 | 0.9 | 478 | 1 | NRFA_ECOLI         |
| 99  | 7 | 0.9 | 481 | 1 | MUCM_ICTPU         |
| 100 | 7 | 0.9 | 484 | 1 | KPYK_CHLFPN        |
| 101 | 7 | 0.9 | 485 | 1 | KPYK_CHLTR         |
| 102 | 7 | 0.9 | 488 | 1 | PGD_TREPA          |
| 103 | 7 | 0.9 | 488 | 1 | NU2M_OENBE         |
| 104 | 7 | 0.9 | 489 | 1 | NU2M_WARPO         |
| 105 | 7 | 0.9 | 497 | 1 | C711_ARATH         |
| 106 | 7 | 0.9 | 504 | 1 | DEGP_RHIME         |

|        |              |
|--------|--------------|
| Q09926 | schizosacch  |
| P73935 | synechocyst  |
| P34463 | caenorhabdi  |
| P15870 | strongyloce  |
| P00873 | chlamydomon  |
| P08475 | chlamydomon  |
| P49209 | arabidopsais |
| Q9hm55 | thermoplasm  |
| Q8Q0f2 | methanosarc  |
| P04892 | bacterioph   |
| P15864 | mus musculu  |
| P15865 | rattus norv  |
| P10412 | homo sapien  |
| P43274 | mus musculu  |
| P43277 | mus musculu  |
| Q03662 | nicotiana t  |
| Q03663 | nicotiana t  |
| Q64757 | avian adeno  |
| Q98m18 | thizobium l  |
| O59827 | schizosacch  |
| O06737 | bacillus su  |
| Q15546 | homo sapien  |
| P02745 | homo sapien  |
| Q58747 | methanococc  |
| Q00730 | candida alb  |
| Q00123 | ictaluriid h |
| P31953 | mycobacteri  |
| P53731 | saccharomyc  |
| Q0214  | saccharomyc  |
| P17676 | homo sapien  |
| O02755 | bos taurus   |
| Q8kd01 | chlorobium   |
| Q9hs49 | halobacteri  |
| Q10175 | schizosacch  |
| P06790 | human papil  |
| Q8u9w3 | agrobacteri  |
| P36794 | human papil  |
| Q87mk5 | humio para   |
| Q27180 | pyrococcus   |
| Q8u086 | pyrococcus   |
| P27131 | naegleria f  |
| Q05214 | nicotiana t  |
| P30167 | solanum tub  |
| P33494 | arabidopsais |
| P30169 | solanum tub  |
| P07676 | eschericia   |
| P35713 | homo sapien  |
| P21679 | bacterioph   |
| P30872 | homo sapien  |
| P30873 | mus musculu  |
| P28646 | rattus norv  |
| Q00366 | mesocricetu  |
| Q09669 | schizosacch  |
| Q8pnu4 | xanthomonas  |
| P42808 | xanthomonas  |
| Q43058 | physcomitre  |
| Q8pc67 | xanthomonas  |
| Q9pbc7 | ureaplasma   |
| Q9pjb0 | campylobact  |
| Q971s5 | clostridium  |
| P27657 | rattus norv  |
| Q01656 | chlamydomon  |
| P50579 | homo sapien  |
| O08663 | mus musculu  |
| P32050 | eschericia   |
| P23735 | ictalurus p  |
| Q9z984 | chlamydia p  |
| P94685 | chlamydia t  |
| O83351 | treponema p  |
| P93401 | oenothera b  |
| P26846 | marichantia  |
| Q9s8b6 | arabidopsais |
| Q52894 | rhizobium m  |

|     |   |     |      |   |            |        |              |     |   |     |      |   |             |        |              |
|-----|---|-----|------|---|------------|--------|--------------|-----|---|-----|------|---|-------------|--------|--------------|
| 107 | 7 | 0.9 | 507  | 1 | HSK1_SCHPO | P50582 | schizosacch  | 180 | 7 | 0.9 | 2434 | 1 | YCF1_OENHO  | Q9mth5 | oenothera h  |
| 108 | 7 | 0.9 | 511  | 1 | HRD2_SIRCO | P18183 | streptomycse | 181 | 7 | 0.9 | 2561 | 1 | PPS1_BACSU  | P39845 | bacillus su  |
| 109 | 7 | 0.9 | 515  | 1 | NU2M_BETVU | P15688 | beta vulgar  | 182 | 7 | 0.9 | 3131 | 1 | ESV1_FUSEBQ | Q00869 | fusarium eq  |
| 110 | 7 | 0.9 | 522  | 1 | KAPR_USTMA | P49605 | ustilago ma  | 183 | 6 | 0.8 | 29   | 1 | DMD_RAT     | P11530 | rattus norv  |
| 111 | 7 | 0.9 | 526  | 1 | KPKK_ASPNG | Q12669 | aspergillus  | 184 | 6 | 0.8 | 38   | 1 | PSBM_CYAPA  | P48107 | cyanophora   |
| 112 | 7 | 0.9 | 529  | 1 | SWA_DROPS  | Q9u915 | drosophila   | 185 | 6 | 0.8 | 40   | 1 | Y382_TREFA  | O83397 | treponema p  |
| 113 | 7 | 0.9 | 535  | 1 | THS6_SULSH | Q9h121 | sulfolobus   | 186 | 6 | 0.8 | 50   | 1 | LAH6_RHOAC  | P53092 | rhodospseudo |
| 114 | 7 | 0.9 | 535  | 1 | THS6_SULSO | Q9v217 | sulfolobus   | 187 | 6 | 0.8 | 51   | 1 | ATPE_MOUSE  | P56382 | mus musculus |
| 115 | 7 | 0.9 | 540  | 1 | ASD1_NEUCR | P78710 | neurospora   | 188 | 6 | 0.8 | 58   | 1 | DP3B_MYCBO  | P80996 | mycobacteri  |
| 116 | 7 | 0.9 | 540  | 1 | FTAB_KUEFN | Q9a947 | klebsiella   | 189 | 6 | 0.8 | 62   | 1 | GVPA_OSCAG  | P82025 | gonatodes v  |
| 117 | 7 | 0.9 | 543  | 1 | EDS5_ARATH | Q94510 | arabidopsis  | 190 | 6 | 0.8 | 62   | 1 | LOTA_GONVI  | P82045 | locusta mig  |
| 118 | 7 | 0.9 | 547  | 1 | SPAK_HUMAN | Q9u68  | homo sapien  | 191 | 6 | 0.8 | 65   | 1 | ONP1_FOCMI  | P82045 | locusta mig  |
| 119 | 7 | 0.9 | 549  | 1 | LIP1_CANRU | P20261 | candida rug  | 192 | 6 | 0.8 | 66   | 1 | LAH2_RHOA   | P82045 | locusta mig  |
| 120 | 7 | 0.9 | 549  | 1 | LIP3_CANRU | P32947 | candida rug  | 193 | 6 | 0.8 | 66   | 1 | MOR2_BOMMO  | P82045 | locusta mig  |
| 121 | 7 | 0.9 | 549  | 1 | LIP5_CANRU | P32949 | candida rug  | 194 | 6 | 0.8 | 66   | 1 | MOR2_BOMMO  | P82045 | locusta mig  |
| 122 | 7 | 0.9 | 553  | 1 | SPAK_RAT   | O88506 | rattus norv  | 195 | 6 | 0.8 | 70   | 1 | GVPA_ANAFL  | P10397 | anabaena fl  |
| 123 | 7 | 0.9 | 572  | 1 | CES6_MOUSE | O88506 | rattus norv  | 196 | 6 | 0.8 | 70   | 1 | GVPA_ANAFL  | P10397 | anabaena fl  |
| 124 | 7 | 0.9 | 572  | 1 | DPY2_BOVIN | O88506 | rattus norv  | 197 | 6 | 0.8 | 70   | 1 | GVPA_ANAFL  | P10397 | anabaena fl  |
| 125 | 7 | 0.9 | 572  | 1 | DPY2_CHICK | O88506 | rattus norv  | 198 | 6 | 0.8 | 70   | 1 | GVPA_ANAFL  | P10397 | anabaena fl  |
| 126 | 7 | 0.9 | 572  | 1 | DPY2_HUMAN | Q16555 | homo sapien  | 199 | 6 | 0.8 | 70   | 1 | GVPA_ANAFL  | P10397 | anabaena fl  |
| 127 | 7 | 0.9 | 572  | 1 | DPY2_MOUSE | O88553 | mus musculus | 200 | 6 | 0.8 | 70   | 1 | GVPA_ANAFL  | P10397 | anabaena fl  |
| 128 | 7 | 0.9 | 572  | 1 | DPY2_RAT   | P47942 | rattus norv  | 201 | 6 | 0.8 | 70   | 1 | GVPA_ANAFL  | P10397 | anabaena fl  |
| 129 | 7 | 0.9 | 575  | 1 | NRD1_YEAST | P53617 | saccharomyc  | 202 | 6 | 0.8 | 80   | 1 | ACP_OOSI    | Q8xful | anabaena sp  |
| 130 | 7 | 0.9 | 578  | 1 | CES6_HUMAN | O9bxc6 | homo sapien  | 203 | 6 | 0.8 | 80   | 1 | YTM6_BACSU  | P07060 | flavivella d |
| 131 | 7 | 0.9 | 595  | 1 | ZN85_HUMAN | Q03923 | homo sapien  | 204 | 6 | 0.8 | 81   | 1 | EX7S_PSESM  | O9r3v0 | planktothri  |
| 132 | 7 | 0.9 | 603  | 1 | FTSH_CYAME | Q9c183 | cyanditosch  | 205 | 6 | 0.8 | 81   | 1 | OGA2_SALTY  | O13907 | cyanidium c  |
| 133 | 7 | 0.9 | 610  | 1 | RYK1_DROME | Q27324 | drosophila   | 206 | 6 | 0.8 | 82   | 1 | ANPA_PSEAM  | P49167 | saccharomyc  |
| 134 | 7 | 0.9 | 614  | 1 | FTSH_CYACA | O19922 | cyanditosch  | 207 | 6 | 0.8 | 82   | 1 | ASR_SALTY   | P22402 | streptomyc   |
| 135 | 7 | 0.9 | 627  | 1 | FTSH_SYNY3 | Q55700 | synchocyst   | 208 | 6 | 0.8 | 83   | 1 | ASR_SALTY   | Q34365 | bacillus su  |
| 136 | 7 | 0.9 | 628  | 1 | FTSH_PORPU | P51327 | porphyra pu  | 209 | 6 | 0.8 | 84   | 1 | YCEP_ECOLI  | O889p9 | pseudomonas  |
| 137 | 7 | 0.9 | 628  | 1 | GIDA_WIGBR | Q8d3k0 | wiggleswort  | 210 | 6 | 0.8 | 84   | 1 | YCEP_ECOLI  | P75927 | escherichia  |
| 138 | 7 | 0.9 | 629  | 1 | QDP2_ECOLI | P06959 | escherichia  | 211 | 6 | 0.8 | 85   | 1 | WIRB_WHEAT  | Q01481 | tritium ae   |
| 139 | 7 | 0.9 | 631  | 1 | FTSH_GUTH  | O78516 | guillardia   | 212 | 6 | 0.8 | 85   | 1 | WIRB_WHEAT  | Q30089 | archaeoglob  |
| 140 | 7 | 0.9 | 646  | 1 | LEM3_BOVIN | P42201 | bos taurus   | 213 | 6 | 0.8 | 88   | 1 | WIRA_WHEAT  | Q01482 | tritium ae   |
| 141 | 7 | 0.9 | 655  | 1 | ACDV_RAT   | P45953 | rattus norv  | 214 | 6 | 0.8 | 89   | 1 | YIHD_SCOLI  | P32126 | escherichia  |
| 142 | 7 | 0.9 | 656  | 1 | ACDV_MOUSE | P50544 | mus musculus | 215 | 6 | 0.8 | 89   | 1 | YIHD_SCOLI  | Q02267 | human papil  |
| 143 | 7 | 0.9 | 659  | 1 | VEB1_FPV03 | P36719 | human papil  | 216 | 6 | 0.8 | 91   | 1 | Y022_ANAFU  | Q30213 | archaeoglob  |
| 144 | 7 | 0.9 | 668  | 1 | YBFI_YEAST | P34217 | saccharomyc  | 217 | 6 | 0.8 | 92   | 1 | PAK2_ANAFU  | Q8y1k8 | anabaena sp  |
| 145 | 7 | 0.9 | 669  | 1 | TSEA_HUMAN | O8u566 | homo sapien  | 218 | 6 | 0.8 | 92   | 1 | RS19_WIGBR  | Q8d208 | wiggleswort  |
| 146 | 7 | 0.9 | 672  | 1 | BGAL_BACST | P19658 | bacillus st  | 219 | 6 | 0.8 | 92   | 1 | VG80_BPML5  | Q05295 | mycobacteri  |
| 147 | 7 | 0.9 | 696  | 1 | SYM_XANCP  | Q8pay7 | xanthomonas  | 220 | 6 | 0.8 | 96   | 1 | RS6_MYCTU   | P71710 | mycobacteri  |
| 148 | 7 | 0.9 | 732  | 1 | DD35_CAEEL | Q9bkd8 | caenorhabdi  | 221 | 6 | 0.8 | 97   | 1 | ANP_LINFE   | P05031 | limanda fer  |
| 149 | 7 | 0.9 | 747  | 1 | NU3C_LOTJA | Q9bpb6 | lotus japon  | 222 | 6 | 0.8 | 98   | 1 | NULM_SCYCA  | Q79409 | scyllorhinu  |
| 150 | 7 | 0.9 | 752  | 1 | R511_TRYCR | P18269 | trypanosoma  | 223 | 6 | 0.8 | 100  | 1 | Y134_MYCPN  | P75502 | mycoplasma   |
| 151 | 7 | 0.9 | 754  | 1 | RIR1_BPT4  | P32282 | bacterioph   | 224 | 6 | 0.8 | 101  | 1 | ASR_ENTCL   | Q93mh6 | enterobacte  |
| 152 | 7 | 0.9 | 769  | 1 | LEM3_SHEEP | P98109 | ovis aries   | 225 | 6 | 0.8 | 101  | 1 | Y127_UREPA  | Q9p177 | ureaplasma   |
| 153 | 7 | 0.9 | 782  | 1 | YMY2_YEAST | P40157 | saccharomyc  | 226 | 6 | 0.8 | 103  | 1 | NIEC_PARDE  | Q51702 | paracoccus   |
| 154 | 7 | 0.9 | 830  | 1 | VBP3_HUMAN | Q13488 | h. vacuolar  | 227 | 6 | 0.8 | 104  | 1 | LEI1_HELAN  | P02253 | bos taurus   |
| 155 | 7 | 0.9 | 837  | 1 | HFC1_HAEIN | P33397 | haemophilus  | 228 | 6 | 0.8 | 104  | 1 | LEI1_HELAN  | P48515 | heilaathus   |
| 156 | 7 | 0.9 | 837  | 1 | HFC2_HAEIN | P45997 | haemophilus  | 229 | 6 | 0.8 | 104  | 1 | YO32_HALNI  | Q9hmq5 | haelobacteri |
| 157 | 7 | 0.9 | 837  | 1 | HFC3_HAEIN | P45998 | haemophilus  | 230 | 6 | 0.8 | 105  | 1 | RLA2_DICDI  | P22683 | dictyosteli  |
| 158 | 7 | 0.9 | 849  | 1 | PHSG_SYNY3 | P73511 | synchocyst   | 231 | 6 | 0.8 | 108  | 1 | CYC_NEUCR   | P00048 | neurospora   |
| 159 | 7 | 0.9 | 853  | 1 | PHS1_DICDI | Q00766 | dictyosteli  | 232 | 6 | 0.8 | 108  | 1 | CYC_STELP   | Q96vp3 | curvularia   |
| 160 | 7 | 0.9 | 878  | 1 | YBM5_SCHPO | Q10337 | schizosacch  | 233 | 6 | 0.8 | 108  | 1 | CYC_STELP   | Q41346 | stellaria l  |
| 161 | 7 | 0.9 | 904  | 1 | YGLB_HSV23 | P06763 | herpes simp  | 234 | 6 | 0.8 | 108  | 1 | IATP_PIG    | Q29307 | sus scrofa   |
| 162 | 7 | 0.9 | 904  | 1 | YGLB_HSV2H | P06666 | herpes simp  | 235 | 6 | 0.8 | 108  | 1 | IATP_PIG    | P56181 | homo sapien  |
| 163 | 7 | 0.9 | 916  | 1 | PNAL_AJECA | Q07421 | ajellomyces  | 236 | 6 | 0.8 | 109  | 1 | CYC_DEBHA   | P00043 | debaromyce   |
| 164 | 7 | 0.9 | 919  | 1 | PNAL_SCHPO | P09627 | schizosacch  | 237 | 6 | 0.8 | 109  | 1 | CYC_DEBEC   | P19681 | debaromyce   |
| 165 | 7 | 0.9 | 920  | 1 | PNAL_NEUCR | P07038 | neurospora   | 238 | 6 | 0.8 | 109  | 1 | RL1_AQUYP   | Q9xdl1 | aquifex pyr  |
| 166 | 7 | 0.9 | 925  | 1 | GLHR_HUMAN | P35409 | anthopleura  | 239 | 6 | 0.8 | 110  | 1 | RL37_TETTH  | P24002 | tetrahymena  |
| 167 | 7 | 0.9 | 930  | 1 | SM6C_HUMAN | Q9h3t2 | homo sapien  | 240 | 6 | 0.8 | 110  | 1 | KDGF_BRWCH  | Q05527 | erwinia chr  |
| 168 | 7 | 0.9 | 983  | 1 | ESYN_GIBPU | Q00868 | gibberella   | 241 | 6 | 0.8 | 110  | 1 | RLA1_ARATH  | Q23095 | arabidopsis  |
| 169 | 7 | 0.9 | 998  | 1 | LONH_PYRAB | Q9uyce | pyrococcus   | 242 | 6 | 0.8 | 110  | 1 | SECG_ECOLI  | P33582 | escherichia  |
| 170 | 7 | 0.9 | 1010 | 1 | PNAL_SCHPO | P28876 | schizosacch  | 243 | 6 | 0.8 | 111  | 1 | SECG_ECOLI  | P00047 | thermonyces  |
| 171 | 7 | 0.9 | 1053 | 1 | SAL4_HUMAN | Q9ujq4 | homo sapien  | 244 | 6 | 0.8 | 111  | 1 | YS80_XANCP  | Q8p6t5 | xanthomonas  |
| 172 | 7 | 0.9 | 1127 | 1 | Y855_TREPA | Q93827 | treponema p  | 245 | 6 | 0.8 | 112  | 1 | CYC1_ARATH  | P29380 | arabidopsis  |
| 173 | 7 | 0.9 | 1147 | 1 | MYSB_ACACA | P19706 | acanthamoeb  | 246 | 6 | 0.8 | 112  | 1 | HMGD_DROME  | Q05783 | synchocyst   |
| 174 | 7 | 0.9 | 1164 | 1 | KEL1_YEAST | P38853 | saccharomyc  | 247 | 6 | 0.8 | 112  | 1 | PSBW_SYNY3  | Q55356 | synchocyst   |
| 175 | 7 | 0.9 | 1387 | 1 | TROP_HUMAN | Q12816 | homo sapien  | 248 | 6 | 0.8 | 113  | 1 | HYPA_RHOSH  | Q9rem0 | rhodobacter  |
| 176 | 7 | 0.9 | 1477 | 1 | XELC_DROME | Q04652 | drosophila   | 249 | 6 | 0.8 | 114  | 1 | CPG1_CANPG  | P81580 | cancer pagu  |
| 177 | 7 | 0.9 | 1567 | 1 | ICEN_XANCT | P8127  | xanthomonas  | 250 | 6 | 0.8 | 114  | 1 | ETS_RABIT   | P19398 | oryctolagus  |
| 178 | 7 | 0.9 | 1573 | 1 | AROI_BACSU | Q9p7r0 | s pentafunc  | 251 | 6 | 0.8 | 119  | 1 | RM14_TETPY  | P10850 | tetrahymena  |
| 179 | 7 | 0.9 | 1585 | 1 | YQBO_SCHPO | P45931 | bacillus su  | 252 | 6 | 0.8 | 119  | 1 | YLR3_EBV    | P03202 | epstein-bar  |

|        |              |
|--------|--------------|
| Q9mth5 | oenothera h  |
| P39845 | bacillus su  |
| Q00869 | fusarium eq  |
| P11530 | rattus norv  |
| P48107 | cyanophora   |
| O83397 | treponema p  |
| P53092 | rhodospseudo |
| P56382 | mus musculus |
| O39194 | mycobacteri  |
| P80996 | oscillatori  |
| P82025 | gonatodes v  |
| P82045 | locusta mig  |
| P35102 | rhodospseudo |
| P82818 | bombyx mori  |
| O96059 | bombyx mori  |
| P10397 | anabaena fl  |
| Q8xful | anabaena sp  |
| P07060 | flavivella d |
| O9r3v0 | planktothri  |
| O13907 | cyanidium c  |
| P49167 | saccharomyc  |
| P22402 | streptomyc   |
| Q34365 | bacillus su  |
| P49517 | odontella s  |
| O889p9 | pseudomonas  |
| P58650 | salmonella   |
| P04002 | pseudopleur  |
| Q82pk9 | salmonella   |
| Q93mh4 | salmonella   |
| Q03032 | salmonella   |
| P75927 | escherichia  |
| Q01481 | tritium ae   |
| Q30089 | archaeoglob  |
| Q01482 | tritium ae   |
| P32126 | escherichia  |
| Q02267 | human papil  |
| Q30213 | archaeoglob  |
| Q8y1k8 | anabaena sp  |
| Q8d208 | wiggleswort  |
| Q05295 | mycobacteri  |
| P71710 | mycobacteri  |
| P05031 | limanda fer  |
| Q79409 | scyllorhinu  |
| P75502 | mycoplasma   |
| Q93mh6 | enterobacte  |
| Q9p177 | ureaplasma   |
| Q51702 | paracoccus   |
| P02253 | bos taurus   |
| P48515 | heilaathus   |
| Q9hmq5 | haelobacteri |
| P22683 | dictyosteli  |
| P00048 | neurospora   |
| Q96vp3 | curvularia   |
| Q41346 | stellaria l  |
| Q29307 | sus scrofa   |
| P56181 | homo sapien  |
| P00043 | debaromyce   |
| P19681 | debaromyce   |
| Q9xdl1 | aquifex pyr  |
| P24002 | tetrahymena  |
| Q05527 | erwinia chr  |
| Q23095 | arabidopsis  |
| P33582 | escherichia  |
| P00047 | thermonyces  |
| Q8p6t5 | xanthomonas  |
| P29380 | arabidopsis  |
| Q05783 | synchocyst   |
| Q55356 | synchocyst   |
| Q9rem0 | rhodobacter  |
| P81580 | cancer pagu  |
| P19398 | oryctolagus  |
| P10850 | tetrahymena  |
| P03202 | epstein-bar  |

|     |   |            |     |     |   |                    |     |   |            |     |     |   |                     |
|-----|---|------------|-----|-----|---|--------------------|-----|---|------------|-----|-----|---|---------------------|
| 253 | 1 | CUL1 ARADI | 120 | 0.8 | 6 | P80515 araneus dia | 326 | 1 | SSB BRUAB  | 168 | 0.8 | 6 | Q07432 brucella ab  |
| 254 | 1 | ASR2 YERPE | 121 | 0.8 | 6 | Q8dc4 yersinia pe  | 327 | 1 | RLU2 DROME | 169 | 0.8 | 6 | P13238 drosophila   |
| 255 | 1 | YIP9 YEAST | 121 | 0.8 | 6 | P40520 saccharomyc | 328 | 1 | RL10 CHLPN | 170 | 0.8 | 6 | Q929a2 chlamydia p  |
| 256 | 1 | DBH DEIRA  | 122 | 0.8 | 6 | Q9zr89 deinococcus | 329 | 1 | RL17 MYCLE | 170 | 0.8 | 6 | Q8x797 mycobacteri  |
| 257 | 1 | YAB4 AQUAE | 122 | 0.8 | 6 | O6r176 aquifex aeo | 330 | 1 | ACT3 LYTPI | 171 | 0.8 | 6 | Q25379 lytechinus   |
| 258 | 1 | PFDB THEVO | 124 | 0.8 | 6 | Q979c4 thermoplasm | 331 | 1 | ACTM LYTPI | 172 | 0.8 | 6 | Q25381 lytechinus   |
| 259 | 1 | SGP2 CHRVI | 124 | 0.8 | 6 | O52179 chromatium  | 332 | 1 | IN7B MOUSE | 172 | 0.8 | 6 | Q920v7 mus musculom |
| 260 | 1 | RPC9 SCHPO | 125 | 0.8 | 6 | Q09177 schizosacch | 333 | 1 | YHL4 HCMVA | 172 | 0.8 | 6 | P09698 human cytom  |
| 261 | 1 | RBFA TREPA | 125 | 0.8 | 6 | O81860 treponema p | 334 | 1 | YCED ECOLI | 173 | 0.8 | 6 | P14189 escherichia  |
| 262 | 1 | COPC YANCI | 126 | 0.8 | 6 | O56797 xanthomonas | 335 | 1 | YCED SALTY | 173 | 0.8 | 6 | O85137 salmonella   |
| 263 | 1 | CRB2 HALNI | 128 | 0.8 | 6 | Q9hwl1 halobacteri | 336 | 1 | GRPB KLEPN | 174 | 0.8 | 6 | P20725 klebsiella   |
| 264 | 1 | EXBD HAEDU | 129 | 0.8 | 6 | O51809 haemophilus | 337 | 1 | YRKL BACSU | 174 | 0.8 | 6 | P54439 bacillus su  |
| 265 | 1 | RS11 BRAJA | 129 | 0.8 | 6 | P59370 bradyrhizob | 338 | 1 | SNAC STRPR | 175 | 0.8 | 6 | P34994 streptomyce  |
| 266 | 1 | OTOR CHICK | 132 | 0.8 | 6 | Q918p6 gallus gall | 339 | 1 | DEF RHIL0  | 176 | 0.8 | 6 | Q98d52 rhizobium l  |
| 267 | 1 | RR8 SPIMX  | 132 | 0.8 | 6 | O98458 spirogyra m | 340 | 1 | HEX1 NEUCR | 176 | 0.8 | 6 | P87252 neurospora   |
| 268 | 1 | ATPE OOSI  | 133 | 0.8 | 6 | P19648 odontella s | 341 | 1 | SCP PATVE  | 176 | 0.8 | 6 | P02637 patinopecte  |
| 269 | 1 | LECA ARTIN | 133 | 0.8 | 6 | P18670 artocarpus  | 342 | 1 | VP26 EBV   | 176 | 0.8 | 6 | P14348 epstein-bar  |
| 270 | 1 | RP9B HUMAN | 133 | 0.8 | 6 | Q9y280 homo sapien | 343 | 1 | BFT ARATH  | 177 | 0.8 | 6 | Q57442 arabidopsis  |
| 271 | 1 | RSB5 SULSO | 133 | 0.8 | 6 | Q980w3 sulfolobus  | 344 | 1 | FLIL BORBU | 177 | 0.8 | 6 | Q57442 arabidopsis  |
| 272 | 1 | RS12 SYNEL | 134 | 0.8 | 6 | P59168 synecococc  | 345 | 1 | IECI ERYVA | 178 | 0.8 | 6 | P34952 erythrina v  |
| 273 | 1 | YBCQ HCMVA | 134 | 0.8 | 6 | P16836 human cytom | 346 | 1 | YPOL THETH | 179 | 0.8 | 6 | P32438 thermus the  |
| 274 | 1 | ULB6 ECOLI | 136 | 0.8 | 6 | P75773 escherichia | 347 | 1 | XJQ0 YEAST | 180 | 0.8 | 6 | P46999 saccharomyc  |
| 275 | 1 | VE6 BPV2   | 137 | 0.8 | 6 | P11302 bovine papi | 348 | 1 | LPQE MYCTU | 182 | 0.8 | 6 | O53569 mycobacteri  |
| 276 | 1 | VIT STRPU  | 137 | 0.8 | 6 | P19615 strongyloce | 349 | 1 | V159 URSPA | 182 | 0.8 | 6 | O9pqv5 ureaplasma   |
| 277 | 1 | YGAV YEAST | 137 | 0.8 | 6 | P50084 saccharomyc | 350 | 1 | PYRE METCA | 183 | 0.8 | 6 | O8vrl1 methylococc  |
| 278 | 1 | VIVA BFT7  | 138 | 0.8 | 6 | P03723 bacterioph  | 351 | 1 | Y098 MYCTU | 183 | 0.8 | 6 | Q10894 mycobacteri  |
| 279 | 1 | ADFX ARATH | 140 | 0.8 | 6 | Q91q81 arabidopsis | 352 | 1 | Y4LS RHISN | 183 | 0.8 | 6 | P55559 rhizobium s  |
| 280 | 1 | SSB DROME  | 140 | 0.8 | 6 | P54622 drosophila  | 353 | 1 | LBP RENRE  | 184 | 0.8 | 6 | P05938 renilla ren  |
| 281 | 1 | VE6 HPV24  | 140 | 0.8 | 6 | P50777 human papil | 354 | 1 | Y583 METTH | 184 | 0.8 | 6 | O26683 methanobact  |
| 282 | 1 | YJ14 YEAST | 140 | 0.8 | 6 | P47111 saccharomyc | 355 | 1 | LEC VICVI  | 185 | 0.8 | 6 | P56625 vicia villo  |
| 283 | 1 | ATPE MYCPU | 141 | 0.8 | 6 | Q98qu6 mycoplasma  | 356 | 1 | MLEI MUGCA | 186 | 0.8 | 6 | P82159 mugil capit  |
| 284 | 1 | CH16 DROGR | 142 | 0.8 | 6 | P24510 drosophila  | 357 | 1 | DEF1 ANASP | 187 | 0.8 | 6 | O8y8k6 anabaena sp  |
| 285 | 1 | NIKE HALNI | 142 | 0.8 | 6 | Q9hn7 halobacteri  | 358 | 1 | DEF FREDI  | 187 | 0.8 | 6 | P4601 firemella d   |
| 286 | 1 | HMBG TETH  | 143 | 0.8 | 6 | P40626 tetrahymena | 359 | 1 | Y418 VIBCH | 187 | 0.8 | 6 | Q9kuu7 vibrio chol  |
| 287 | 1 | NB6M BOVIN | 143 | 0.8 | 6 | Q95kv7 bos taurus  | 360 | 1 | KITH STRGC | 191 | 0.8 | 6 | Q47848 streptococc  |
| 288 | 1 | PSGD HUMAN | 143 | 0.8 | 6 | Q9uc72 homo sapien | 361 | 1 | RUVA THETH | 191 | 0.8 | 6 | Q9f1q3 thermus the  |
| 289 | 1 | FAB4 CAEEL | 145 | 0.8 | 6 | Q23092 caenorhabdi | 362 | 1 | Y893 DROME | 192 | 0.8 | 6 | Q9vv43 drosophila   |
| 290 | 1 | PSE2 ARATH | 145 | 0.8 | 6 | Q9s714 arabidopsis | 363 | 1 | LUXJ VIBFI | 193 | 0.8 | 6 | P35328 vibrio fisc  |
| 291 | 1 | SSIK STRPT | 145 | 0.8 | 6 | O33702 streptomyce | 364 | 1 | RL9 BEA    | 193 | 0.8 | 6 | P30707 pisum sativ  |
| 292 | 1 | ZEPT IRV6  | 145 | 0.8 | 6 | P40629 chilo iride | 365 | 1 | FGF4 CHICK | 194 | 0.8 | 6 | P48804 gallus gall  |
| 293 | 1 | EXBD HAEIN | 147 | 0.8 | 6 | P43009 haemophilus | 366 | 1 | HL SALTR   | 194 | 0.8 | 6 | P02254 salmo trutt  |
| 294 | 1 | NUCA BACSU | 147 | 0.8 | 6 | P12667 bacillus su | 367 | 1 | YB6A SCHPO | 194 | 0.8 | 6 | O14256 schizosacch  |
| 295 | 1 | DTD OCRIH  | 148 | 0.8 | 6 | Q8epre oceanobacil | 368 | 1 | YI32 AGRT7 | 194 | 0.8 | 6 | P05679 agrobacteri  |
| 296 | 1 | YEAL ECOLI | 148 | 0.8 | 6 | P76240 escherichia | 369 | 1 | BTUR ECOLI | 196 | 0.8 | 6 | P13040 escherichia  |
| 297 | 1 | DTD CLOPE  | 149 | 0.8 | 6 | Q8x124 clostridium | 370 | 1 | BTUR SALTY | 196 | 0.8 | 6 | P31570 salmonella   |
| 298 | 1 | MOAC SULSO | 149 | 0.8 | 6 | Q9uxf7 sulfolobus  | 371 | 1 | YBAO ECOLI | 196 | 0.8 | 6 | P45466 escherichia  |
| 299 | 1 | VE6 HPV58  | 149 | 0.8 | 6 | P26555 human papil | 372 | 1 | ISPZ PSRPK | 197 | 0.8 | 6 | P59364 pseudomonas  |
| 300 | 1 | YHS9 CLOPE | 149 | 0.8 | 6 | Q8xj17 clostridium | 373 | 1 | RUVA CLOHI | 197 | 0.8 | 6 | Q9znj6 clostridium  |
| 301 | 1 | ALL5 HEVBR | 150 | 0.8 | 6 | Q39967 hevea bras  | 374 | 1 | COP CLOPE  | 198 | 0.8 | 6 | Q60327 methanococc  |
| 302 | 1 | MOAE CAUCR | 150 | 0.8 | 6 | Q9ac50 caulobacter | 375 | 1 | Y011 METJA | 198 | 0.8 | 6 | P18015 clostridium  |
| 303 | 1 | RL15 MYCGE | 151 | 0.8 | 6 | P47415 mycoplasma  | 376 | 1 | PEMT HUMAN | 198 | 0.8 | 6 | Q9ubm1 homo sapien  |
| 304 | 1 | RL15 MYCPN | 151 | 0.8 | 6 | Q50300 mycoplasma  | 377 | 1 | SDC4 HUMAN | 198 | 0.8 | 6 | P31431 homo sapien  |
| 305 | 1 | ACT4 LYTPI | 154 | 0.8 | 6 | Q25380 lytechinus  | 378 | 1 | NORE YERPE | 199 | 0.8 | 6 | Q9zc45 yersinia pe  |
| 306 | 1 | YGG9 YEAST | 154 | 0.8 | 6 | P53162 saccharomyc | 379 | 1 | COAE BRUNE | 200 | 0.8 | 6 | Q8ye21 brucella me  |
| 307 | 1 | GREX CAUCR | 157 | 0.8 | 6 | Q9a413 caulobacter | 380 | 1 | PCP FYRAB  | 200 | 0.8 | 6 | Q9uyq9 pyrococcus   |
| 308 | 1 | GREX RHIL0 | 157 | 0.8 | 6 | Q98149 rhizobium l | 381 | 1 | NING BP933 | 201 | 0.8 | 6 | Q9zwx1 bacterioph   |
| 309 | 1 | RL24 HUMAN | 157 | 0.8 | 6 | P38563 homo sapien | 382 | 1 | NING BPH19 | 201 | 0.8 | 6 | O48427 bacterioph   |
| 310 | 1 | RS11 DUNTE | 157 | 0.8 | 6 | P42756 dunaliella  | 383 | 1 | CLP1 AGRT5 | 202 | 0.8 | 6 | Q8uek6 agrobacteri  |
| 311 | 1 | YQAA HAEIN | 158 | 0.8 | 6 | P44005 haemophilus | 384 | 1 | NING BPP22 | 203 | 0.8 | 6 | Q38667 bacterioph   |
| 312 | 1 | RL29 HUMAN | 158 | 0.8 | 6 | P47914 homo sapien | 385 | 1 | SAS2 DICI1 | 203 | 0.8 | 6 | P20791 dictyosteli  |
| 313 | 1 | HTF4 PAPH  | 160 | 0.8 | 6 | Q28772 papio hamad | 386 | 1 | HNGT ONCMY | 204 | 0.8 | 6 | P07746 oncorhynch   |
| 314 | 1 | PRSA ECOLI | 161 | 0.8 | 6 | P42184 escherichia | 387 | 1 | NING BPP21 | 204 | 0.8 | 6 | Q9xjq4 bacterioph   |
| 315 | 1 | Y903 XILFA | 161 | 0.8 | 6 | Q9pex3 xyella fas  | 388 | 1 | NING LAMB  | 204 | 0.8 | 6 | P03770 bacterioph   |
| 316 | 1 | CAV2 HUMAN | 162 | 0.8 | 6 | P51836 homo sapien | 389 | 1 | COAE HAEIN | 206 | 0.8 | 6 | P44920 haemophilus  |
| 317 | 1 | NOS2 MACMU | 162 | 0.8 | 6 | O46660 macaca mula | 390 | 1 | DTM PARPN  | 206 | 0.8 | 6 | Q63347 paracoccu    |
| 318 | 1 | DTD NEIMA  | 163 | 0.8 | 6 | Q9j8t7 neisseria m | 391 | 1 | H1 ONCMY   | 206 | 0.8 | 6 | P06350 oncorhynch   |
| 319 | 1 | GREX HELPY | 164 | 0.8 | 6 | Q9k143 neisseria m | 392 | 1 | PCP PYRHO  | 206 | 0.8 | 6 | O58321 pyrococcus   |
| 320 | 1 | GREX HELPY | 164 | 0.8 | 6 | P55978 helicobacte | 393 | 1 | RS4 BUCAP  | 206 | 0.8 | 6 | P41186 buchera ap   |
| 321 | 1 | USC7 CAEEL | 164 | 0.8 | 6 | P34477 caenorhabdi | 394 | 1 | CSF3 HUMAN | 207 | 0.8 | 6 | P09919 homo sapien  |
| 322 | 1 | Y338 STRPY | 164 | 0.8 | 6 | Q9aid3 streptococc | 395 | 1 | ICW3 PSOTE | 207 | 0.8 | 6 | P10822 psophocarpu  |
| 323 | 1 | LE13 GOSHI | 165 | 0.8 | 6 | P09441 gossypium h | 396 | 1 | SCOB HELPY | 207 | 0.8 | 6 | Q9z1e4 helicobacte  |
| 324 | 1 | ET3 RAT    | 167 | 0.8 | 6 | P13207 rattus norv | 397 | 1 | SCOB HELPY | 207 | 0.8 | 6 | P56007 helicobacte  |
| 325 | 1 | RBS SACHY  | 168 | 0.8 | 6 | Q41373 saccharum h | 398 | 1 | NGN1 BRARE | 208 | 0.8 | 6 | O42606 brachydanio  |

|     |   |     |     |   |             |        |              |     |   |     |     |   |             |        |              |
|-----|---|-----|-----|---|-------------|--------|--------------|-----|---|-----|-----|---|-------------|--------|--------------|
| 399 | 6 | 0.8 | 208 | 1 | Y522_CABEL  | Q09364 | caenorhabdi  | 472 | 6 | 0.8 | 231 | 1 | CUT2_CABEL  | P34682 | caenorhabdi  |
| 400 | 6 | 0.8 | 209 | 1 | LIPB_RICCN  | Q92fx0 | rickettsia   | 473 | 6 | 0.8 | 231 | 1 | RNC_LACLA   | Q8chd0 | lactococcus  |
| 401 | 6 | 0.8 | 209 | 1 | PGSA_MYCTU  | Q50611 | mycobacteri  | 474 | 6 | 0.8 | 232 | 1 | TRPF_LIPST  | Q01128 | lipomyces s  |
| 402 | 6 | 0.8 | 210 | 1 | APSI1_SCHPO | Q09790 | schizosacch  | 475 | 6 | 0.8 | 232 | 1 | Y529_METUA  | Q57949 | methanococc  |
| 403 | 6 | 0.8 | 210 | 1 | V150_AQUAE  | Q67702 | aquifex aeo  | 476 | 6 | 0.8 | 234 | 1 | A29B_DROME  | Q46197 | drosophila   |
| 404 | 6 | 0.8 | 211 | 1 | DER1_YEAST  | P38307 | saccharomyc  | 477 | 6 | 0.8 | 234 | 1 | GLPF_STRPN  | P32281 | streptococc  |
| 405 | 6 | 0.8 | 211 | 1 | HIT_FIG     | P06348 | sus scrofa   | 478 | 6 | 0.8 | 234 | 1 | VGP8_EBV    | P32224 | epstein-bar  |
| 406 | 6 | 0.8 | 212 | 1 | H1Z_HUMAN   | P16403 | homo sapien  | 479 | 6 | 0.8 | 235 | 1 | ISPD_ECOS7  | Q8x7y4 | escherichia  |
| 407 | 6 | 0.8 | 212 | 1 | PYRE_ECOS7  | Q8xd99 | escherichia  | 480 | 6 | 0.8 | 235 | 1 | ISPD_ECOL6  | Q8feJ5 | escherichia  |
| 408 | 6 | 0.8 | 212 | 1 | PYRE_ECOLI  | P00495 | escherichia  | 481 | 6 | 0.8 | 235 | 1 | ISPD_ECOLI  | Q46893 | escherichia  |
| 409 | 6 | 0.8 | 212 | 1 | PYRE_SALTY  | Q8z2h5 | salmonella   | 482 | 6 | 0.8 | 235 | 1 | PAC2_SCHPO  | Q10294 | schizosacch  |
| 410 | 6 | 0.8 | 212 | 1 | PYRE_SALTY  | P08870 | salmonella   | 483 | 6 | 0.8 | 236 | 1 | B11_FAT     | P50622 | rattus norv  |
| 411 | 6 | 0.8 | 213 | 1 | LIPB_ECOLI  | P30976 | escherichia  | 484 | 6 | 0.8 | 236 | 1 | MOB1_YEAST  | P04084 | saccharomyc  |
| 412 | 6 | 0.8 | 213 | 1 | PYRE_NEIMA  | Q9ir25 | neisseria m  | 485 | 6 | 0.8 | 236 | 1 | PUR7_PSEAE  | Q914w0 | pseudomonas  |
| 413 | 6 | 0.8 | 213 | 1 | PYRE_PSEAE  | P50587 | pseudomonas  | 486 | 6 | 0.8 | 236 | 1 | Y1HL_ECOLI  | P32133 | escherichia  |
| 414 | 6 | 0.8 | 213 | 1 | PYRE_VIBVU  | Q8ddx5 | vibrio vuln  | 487 | 6 | 0.8 | 236 | 1 | Y1Z6_YEAST  | P47115 | saccharomyc  |
| 415 | 6 | 0.8 | 214 | 1 | ET3_MOUSE   | P48299 | mus musculu  | 488 | 6 | 0.8 | 237 | 1 | B11_HUMAN   | P50621 | homo sapien  |
| 416 | 6 | 0.8 | 214 | 1 | PYRE_PASWU  | Q9cJw4 | pasteurella  | 489 | 6 | 0.8 | 237 | 1 | NGN1_PAROL  | Q91a79 | paralichthy  |
| 417 | 6 | 0.8 | 214 | 1 | PYRE_VIECH  | Q9Kvds | vibrio chol  | 490 | 6 | 0.8 | 237 | 1 | NGN1_HUMAN  | Q92886 | homo sapien  |
| 418 | 6 | 0.8 | 215 | 1 | ERD2_ENTHI  | O44017 | entamoeba h  | 491 | 6 | 0.8 | 237 | 1 | PYRF_FUSNU  | Q87g83 | fusobacteri  |
| 419 | 6 | 0.8 | 215 | 1 | PYRE_SCHPO  | O94331 | schizosacch  | 492 | 6 | 0.8 | 237 | 1 | YHBE_BACSU  | P39132 | bacillus su  |
| 420 | 6 | 0.8 | 215 | 1 | PYRE_YERPE  | Q8zjp7 | yersinia pe  | 493 | 6 | 0.8 | 238 | 1 | PYRE_COCPO  | O93849 | coccidioid   |
| 421 | 6 | 0.8 | 215 | 1 | YCG6_RICCN  | Q92g57 | rickettsia   | 494 | 6 | 0.8 | 238 | 1 | PYRF_ANASP  | P58639 | anabaena sp  |
| 422 | 6 | 0.8 | 216 | 1 | CLPP_ORYSA  | P12209 | oryza sativ  | 495 | 6 | 0.8 | 238 | 1 | RK5_OPOSI   | P49547 | odontella s  |
| 423 | 6 | 0.8 | 217 | 1 | LIPB_PSEAE  | Q9x6v9 | pseudomonas  | 496 | 6 | 0.8 | 238 | 1 | YOXD_BACSU  | P14802 | bacillus su  |
| 424 | 6 | 0.8 | 217 | 1 | R55_METUA   | P34045 | methanococc  | 497 | 6 | 0.8 | 239 | 1 | PRRH_BACHD  | Q9Ka65 | bacillus na  |
| 425 | 6 | 0.8 | 218 | 1 | EST2_PSEFL  | Q93547 | pseudomonas  | 498 | 6 | 0.8 | 239 | 1 | RLPA_NEIMA  | Q9Jsm7 | neisseria m  |
| 426 | 6 | 0.8 | 218 | 1 | H1S1_DEIRA  | Q9ruet | deinococcus  | 499 | 6 | 0.8 | 239 | 1 | YGCN_ECOLI  | P46143 | escherichia  |
| 427 | 6 | 0.8 | 218 | 1 | RP1A_YERPE  | Q8zh8s | yersinia pe  | 500 | 6 | 0.8 | 240 | 1 | BA29_MOUSE  | Q61334 | mus musculu  |
| 428 | 6 | 0.8 | 218 | 1 | SIR_DESVH   | Q05805 | desulfovibr  | 501 | 6 | 0.8 | 240 | 1 | RNC_MYCTU   | Q10962 | mycobacteri  |
| 429 | 6 | 0.8 | 219 | 1 | BAS2_NORV   | Q05175 | rattus norv  | 502 | 6 | 0.8 | 241 | 1 | BUDC_KLETE  | Q04520 | Klebsiella   |
| 430 | 6 | 0.8 | 219 | 1 | PYRE_XANAC  | Q8pf55 | xanthomonas  | 503 | 6 | 0.8 | 241 | 1 | GPS_MSDO    | P46437 | musca domes  |
| 431 | 6 | 0.8 | 219 | 1 | PYRE_XANCP  | Q8p459 | xanthomonas  | 504 | 6 | 0.8 | 242 | 1 | BIDI_HAEIN  | P45209 | haemophilus  |
| 432 | 6 | 0.8 | 219 | 1 | PYRE_YARLI  | P19123 | yarrowia li  | 505 | 6 | 0.8 | 242 | 1 | RECO_ECOLI  | P15027 | escherichia  |
| 433 | 6 | 0.8 | 219 | 1 | RP1A_ECOLI  | P27252 | escherichia  | 506 | 6 | 0.8 | 244 | 1 | MT04_PSEAE  | P916b3 | pseudomonas  |
| 434 | 6 | 0.8 | 219 | 1 | RP1A_ENTMC  | O52398 | edwardsiell  | 507 | 6 | 0.8 | 244 | 1 | NGN1_MOUSE  | P70660 | mus musculu  |
| 435 | 6 | 0.8 | 219 | 1 | RP1A_ECOLI  | Q8rly6 | enterobacte  | 508 | 6 | 0.8 | 244 | 1 | NGN1_RAT    | P70595 | rattus norv  |
| 436 | 6 | 0.8 | 219 | 1 | RP1A_SALTY  | Q8xek1 | salmonella   | 509 | 6 | 0.8 | 244 | 1 | PYRF_BACCL  | P46535 | bacillus ca  |
| 437 | 6 | 0.8 | 219 | 1 | YEGF_ECOLI  | P76396 | escherichia  | 510 | 6 | 0.8 | 244 | 1 | TRUA_CLOAB  | Q97ell | clostridium  |
| 438 | 6 | 0.8 | 220 | 1 | SER2_GALME  | O96615 | galleria me  | 511 | 6 | 0.8 | 245 | 1 | Y399_ARCFU  | O29848 | archaeoglob  |
| 439 | 6 | 0.8 | 220 | 1 | Y4C2_PASMU  | Q9cmn5 | pasteurella  | 512 | 6 | 0.8 | 245 | 1 | H1_MAIZE    | P23444 | zea mays (m  |
| 440 | 6 | 0.8 | 221 | 1 | QSEB_HAEIN  | P45337 | haemophilus  | 513 | 6 | 0.8 | 245 | 1 | RNC_BRUME   | Q8y974 | brucella me  |
| 441 | 6 | 0.8 | 222 | 1 | MDT7_YEAST  | Q88278 | saccharomyc  | 514 | 6 | 0.8 | 246 | 1 | YDS3_SCHPO  | Q04179 | schizosacch  |
| 442 | 6 | 0.8 | 222 | 1 | VV_SVS      | P12207 | simian viru  | 515 | 6 | 0.8 | 246 | 1 | COAT_PASVK  | Q00323 | panicum str  |
| 443 | 6 | 0.8 | 223 | 1 | DEC1_YERPE  | Q8zgh4 | yersinia pe  | 516 | 6 | 0.8 | 247 | 1 | HIS4_AGRTS  | P58790 | agrobacteri  |
| 444 | 6 | 0.8 | 223 | 1 | GLX3_TOBAC  | Q03664 | nicotiana t  | 517 | 6 | 0.8 | 247 | 1 | YABB_BACSU  | P37543 | bacillus su  |
| 445 | 6 | 0.8 | 223 | 1 | MOXP_FARDE  | P29904 | paracoccus   | 518 | 6 | 0.8 | 248 | 1 | MPA_ECOLI   | P77486 | escherichia  |
| 446 | 6 | 0.8 | 224 | 1 | REP1_ZYGFE  | P13741 | zygosacchar  | 519 | 6 | 0.8 | 248 | 1 | PNK_SULTO   | O69yn6 | sulfolobus   |
| 447 | 6 | 0.8 | 224 | 1 | RL1_METTH   | O27716 | methanobact  | 520 | 6 | 0.8 | 248 | 1 | RS3_ACHSP   | P41118 | acholeplasm  |
| 448 | 6 | 0.8 | 224 | 1 | SPF_BACSU   | P39135 | bacillus su  | 521 | 6 | 0.8 | 249 | 1 | KYXG_KLULA  | P09807 | kluyveromyc  |
| 449 | 6 | 0.8 | 224 | 1 | VV_MUMPE    | P30927 | mumps virus  | 522 | 6 | 0.8 | 249 | 1 | LEG3_HUMAN  | P17931 | homo sapien  |
| 450 | 6 | 0.8 | 224 | 1 | VV_MUMPM    | P30948 | mumps virus  | 523 | 6 | 0.8 | 249 | 1 | PNK_ARCFU   | Q30297 | archaeoglob  |
| 451 | 6 | 0.8 | 224 | 1 | VV_MUMPS    | P33463 | mumps virus  | 524 | 6 | 0.8 | 249 | 1 | PSA3_ARATH  | O32715 | arabidopsais |
| 452 | 6 | 0.8 | 225 | 1 | R5S_METVA   | P14036 | methanococc  | 525 | 6 | 0.8 | 249 | 1 | PSA3_ORYSA  | Q91su0 | oryza sativ  |
| 453 | 6 | 0.8 | 226 | 1 | AROD_HALN1  | Q9hsb4 | halobacteri  | 526 | 6 | 0.8 | 249 | 1 | PSA3_SPIOL  | O24362 | spinacia ol  |
| 454 | 6 | 0.8 | 226 | 1 | BASP_BOVIN  | P80724 | bos taurus   | 527 | 6 | 0.8 | 250 | 1 | SBD5_MOUSE  | P70122 | mus musculu  |
| 455 | 6 | 0.8 | 226 | 1 | BASP_HUMAN  | P80723 | homo sapien  | 528 | 6 | 0.8 | 250 | 1 | YHIQ_NEIGO  | P72077 | neisseria g  |
| 456 | 6 | 0.8 | 226 | 1 | PYRE_SALTY  | O8xg90 | salmonella   | 529 | 6 | 0.8 | 251 | 1 | BTE4_MOUSE  | P59334 | mus musculu  |
| 457 | 6 | 0.8 | 226 | 1 | PYRE_KLULA  | O13474 | kluyveromyc  | 530 | 6 | 0.8 | 251 | 1 | RS3_PHYSI   | O66095 | phytoplasm   |
| 458 | 6 | 0.8 | 226 | 1 | PYRE_PALSO  | Q9y342 | ralesconia s | 531 | 6 | 0.8 | 251 | 1 | TPIS_NEIMA  | Q9jw31 | neisseria m  |
| 459 | 6 | 0.8 | 226 | 1 | PYRE_YEAST  | P13298 | saccharomyc  | 532 | 6 | 0.8 | 251 | 1 | TPIS_NEIMEB | Q9jxt8 | neisseria m  |
| 460 | 6 | 0.8 | 226 | 1 | TRPF_METUA  | Q57893 | methanococc  | 533 | 6 | 0.8 | 251 | 1 | Y484_MYCTU  | Q11150 | mycobacteri  |
| 461 | 6 | 0.8 | 226 | 1 | VATE_HETSC  | Q9ul95 | heterodera   | 534 | 6 | 0.8 | 252 | 1 | BTE4_HUMAN  | P95bkl | homo sapien  |
| 462 | 6 | 0.8 | 227 | 1 | THY1_XENLA  | P01152 | xenopus lae  | 535 | 6 | 0.8 | 252 | 1 | PSA3_ACACA  | P90513 | acanthamoeb  |
| 463 | 6 | 0.8 | 227 | 1 | YB61_DROME  | Q9vna4 | drosophila   | 536 | 6 | 0.8 | 253 | 1 | ENVY_ECOLI  | P10805 | escherichia  |
| 464 | 6 | 0.8 | 227 | 1 | YCA5_MYCH   | O50188 | mycoplasma   | 537 | 6 | 0.8 | 253 | 1 | RS3_CHLTE   | P08048 | chlorobium   |
| 465 | 6 | 0.8 | 228 | 1 | API_TRICO   | P59704 | trichostron  | 538 | 6 | 0.8 | 253 | 1 | TPIS_CORPA  | Q8kah8 | chlorobium   |
| 466 | 6 | 0.8 | 229 | 1 | R521_GEOCY  | Q8wgq3 | geodia cydo  | 539 | 6 | 0.8 | 253 | 1 | Y166_DEIRA  | Q91t03 | deinococcus  |
| 467 | 6 | 0.8 | 230 | 1 | PRF7_GAEEL  | Q1083  | caenorhabdi  | 540 | 6 | 0.8 | 253 | 1 | YOM4_CABEL  | Q32202 | caenorhabdi  |
| 468 | 6 | 0.8 | 230 | 1 | PYRF_STAAM  | Q99ur4 | staphylococ  | 541 | 6 | 0.8 | 254 | 1 | AZLC_BACSU  | O07942 | bacillus su  |
| 469 | 6 | 0.8 | 230 | 1 | SGY1_MOUSE  | Q9qz19 | mus musculu  | 542 | 6 | 0.8 | 255 | 1 | GRPS_MYXXA  | P95333 | myxococcus   |
| 470 | 6 | 0.8 | 231 | 1 | TN3_NEILA   | Q51083 | neisseria l  | 543 | 6 | 0.8 | 256 | 1 | TAM_AGRTS   | Q8uh15 | agrobacteri  |
| 471 | 6 | 0.8 | 231 | 1 | ALSE_ECOLI  | P32719 | escherichia  | 544 | 6 | 0.8 | 256 | 1 | Y165_METUA  | Q97629 | methanococc  |

|     |   |            |        |               |     |   |     |     |   |            |        |              |
|-----|---|------------|--------|---------------|-----|---|-----|-----|---|------------|--------|--------------|
| 545 | 1 | MADE_FUSNN | Q8rea7 | fusobacteri   | 618 | 5 | 0.8 | 284 | 1 | IPYR_PICPA | 013505 | pichia past  |
| 546 | 1 | NCAP_TSWV3 | P36293 | tomato spot   | 619 | 6 | 0.8 | 284 | 1 | LICH_VIBCH | 007350 | vibrio chol  |
| 547 | 1 | RL2_ARATH  | P46286 | arabidopsis   | 620 | 6 | 0.8 | 285 | 1 | DAP1_ANASP | P54897 | anabaena sp  |
| 548 | 1 | MSX2_CHICK | P28362 | gallus gall   | 621 | 6 | 0.8 | 285 | 1 | YIF0_YEAST | P40186 | saccharomyc  |
| 549 | 1 | MSX2_COTJA | P23410 | coturnix co   | 622 | 6 | 0.8 | 286 | 1 | FOLD_CHLNP | Q928k3 | chlamydia p  |
| 550 | 1 | CCG6_MOUSE | Q8vhw3 | mus musculu   | 623 | 6 | 0.8 | 286 | 1 | Y320_MYCGB | P47562 | mycoplasma   |
| 551 | 1 | CCG6_RAT   | Q8vhw7 | rattus norv   | 624 | 6 | 0.8 | 286 | 1 | YQ67_STRP3 | Q8K5ul | streptococ   |
| 552 | 1 | TATD_ECOLI | P27859 | escherichia   | 625 | 6 | 0.8 | 286 | 1 | YQ36_STRPY | Q93y04 | streptococ   |
| 553 | 1 | EUTC_BRAJA | Q89qx6 | bradyrhizob   | 626 | 6 | 0.8 | 287 | 1 | AUF_BACST  | P94453 | bacillus st  |
| 554 | 1 | IEO_NPVAC  | P41710 | autographa    | 627 | 6 | 0.8 | 287 | 1 | EUTC_PSESM | Q889m3 | pseudomonas  |
| 555 | 1 | Y4ZC_RHISN | P55730 | rhizobium s   | 628 | 6 | 0.8 | 287 | 1 | Y320_MYCPN | P75328 | mycoplasma   |
| 556 | 1 | Y602_METJA | Q58019 | methanococ    | 629 | 6 | 0.8 | 287 | 1 | YD04_SYNV3 | P73599 | synecocyst   |
| 557 | 1 | COR4_WHEAT | P46524 | triticocum ae | 630 | 6 | 0.8 | 288 | 1 | BSN2_BACSU | Q32150 | bacillus su  |
| 558 | 1 | LPXA_VIBCH | Q9kpw4 | vibrio chol   | 631 | 6 | 0.8 | 288 | 1 | PCD1_HUMAN | Q15116 | homo sapien  |
| 559 | 1 | PSB2_RAT   | P34067 | rattus norv   | 632 | 6 | 0.8 | 289 | 1 | HEM3_METTH | Q28960 | methanobact  |
| 560 | 1 | SPAK_SALTY | P40701 | salmonella    | 633 | 6 | 0.8 | 289 | 1 | ZN75_HUMAN | P51815 | homo sapien  |
| 561 | 1 | TRPA_CIOAB | Q97ef6 | clostridium   | 634 | 6 | 0.8 | 292 | 1 | CORC_ECOLI | P77392 | escherichia  |
| 562 | 1 | ARGB_MORS4 | Q9k4z5 | moritella s   | 635 | 6 | 0.8 | 292 | 1 | CORC_SALTY | Q9r874 | salmonella   |
| 563 | 1 | RS3_FALSO  | Q8xv18 | ralstonia s   | 636 | 6 | 0.8 | 292 | 1 | Y938_XYLPA | Q9peu0 | xyliella fas |
| 564 | 1 | IHA_SHEEP  | P38440 | ovis aries    | 637 | 6 | 0.8 | 293 | 1 | PRTT_TRIAL | P20015 | tritirachiu  |
| 565 | 1 | RS3_ACHLA  | P29223 | acholeplasm   | 638 | 6 | 0.8 | 293 | 1 | RS3_EORBU  | P94273 | borrelia bu  |
| 566 | 1 | ACT_CRAVI  | Q92193 | crassostrea   | 639 | 6 | 0.8 | 293 | 1 | YKUN_BACSU | Q34827 | bacillus su  |
| 567 | 1 | MURI_LISIN | Q32ch2 | listeria in   | 640 | 6 | 0.8 | 294 | 1 | CHIB_TOBAC | P29061 | nicotiana t  |
| 568 | 1 | MURI_LISMO | Q5y7n7 | listeria mo   | 641 | 6 | 0.8 | 294 | 1 | MDH_ARCFU  | Q08349 | archaeoglob  |
| 569 | 1 | DAPB_PSESZ | Q52419 | pseudomonas   | 642 | 6 | 0.8 | 294 | 1 | NUSG_STRGR | P36260 | streptomyce  |
| 570 | 1 | IF2A_HALN1 | Q9hrt8 | halobacteri   | 643 | 6 | 0.8 | 294 | 1 | YQOI_BACSU | P46340 | bacillus su  |
| 571 | 1 | MSX2_HUMAN | P35548 | homo sapien   | 644 | 6 | 0.8 | 295 | 1 | MPR1_SCHPO | Q94321 | schizosacch  |
| 572 | 1 | MSX2_MOUSE | Q03358 | mus musculu   | 645 | 6 | 0.8 | 296 | 1 | CSBB_MOUSE | P28033 | mus musculu  |
| 573 | 1 | RP30_HUMAN | P78346 | homo sapien   | 646 | 6 | 0.8 | 296 | 1 | RL5_CHICK  | P22451 | gallus gall  |
| 574 | 1 | RPNA_YEAST | P38886 | saccharomyc   | 647 | 6 | 0.8 | 297 | 1 | ARGB_RHIL0 | Q98d76 | rhizobium l  |
| 575 | 1 | TR1_DROME  | P36188 | drosophila    | 648 | 6 | 0.8 | 297 | 1 | CEBB_RAT   | P21272 | rattus norv  |
| 576 | 1 | UPK_RHIME  | P928p2 | rhizobium m   | 649 | 6 | 0.8 | 297 | 1 | PRTA_STRGR | P00776 | streptomyce  |
| 577 | 1 | YOR3_LORDV | P54636 | lordsdale v   | 650 | 6 | 0.8 | 298 | 1 | MMSB_PSEAP | P28811 | pseudomonas  |
| 578 | 1 | THID_HELPJ | Q92100 | helicobacte   | 651 | 6 | 0.8 | 299 | 1 | BPHC_PSEPA | P11122 | pseudomonas  |
| 579 | 1 | YSI3_STAAU | O05337 | staphylococ   | 652 | 6 | 0.8 | 299 | 1 | HEM6_XANAC | Q8pf76 | xanthomonas  |
| 580 | 1 | EUTC_RALSO | Q8xur0 | ralstonia s   | 653 | 6 | 0.8 | 299 | 1 | MMS3_MYCTU | Q10390 | mycobacteri  |
| 581 | 1 | L181_CHELU | Q03965 | chlamydomon   | 654 | 6 | 0.8 | 299 | 1 | NUSG_STRVG | P27309 | streptomyce  |
| 582 | 1 | MIA2_SALTY | Q08015 | salmonella    | 655 | 6 | 0.8 | 299 | 1 | RL22_DROME | P50887 | drosophila   |
| 583 | 1 | THID_HELPY | O25515 | helicobacte   | 656 | 6 | 0.8 | 299 | 1 | YXAK_BACSU | P37513 | bacillus su  |
| 584 | 1 | EUTC_XANAC | Q8pk11 | xanthomonas   | 657 | 6 | 0.8 | 300 | 1 | COX2_YEAST | P19516 | saccharomyc  |
| 585 | 1 | FLIP_TREDE | Q9x5a6 | treponema d   | 658 | 6 | 0.8 | 300 | 1 | TRPI_RABIT | P19761 | oryctolagus  |
| 586 | 1 | MURI_LACLA | O9cg28 | lactococcus   | 659 | 6 | 0.8 | 301 | 1 | CCSA_GUITH | P22554 | guillardia   |
| 587 | 1 | PDX4_HUMAN | Q13162 | homo sapien   | 660 | 6 | 0.8 | 301 | 1 | PORI_RHOCA | P31243 | rhodobacter  |
| 588 | 1 | THID_BACSU | P39610 | bacillus su   | 661 | 6 | 0.8 | 301 | 1 | UL49_HSV11 | P10233 | herpes simp  |
| 589 | 1 | TRPA_BUCAP | P42389 | buchnera ap   | 662 | 6 | 0.8 | 302 | 1 | CILB_ECOLI | P77770 | escherichia  |
| 590 | 1 | EUTC_PSEPK | Q8qf2  | pseudomonas   | 663 | 6 | 0.8 | 303 | 1 | CBPB_ASTFL | P04069 | astacus flu  |
| 591 | 1 | EUTC_XANCP | Q8p810 | xanthomonas   | 664 | 6 | 0.8 | 303 | 1 | LP12_SYNV3 | P73572 | synecocyst   |
| 592 | 1 | CBF2_CAMJE | Q46105 | campylobact   | 665 | 6 | 0.8 | 303 | 1 | LPXC_PSEAP | P47202 | pseudomonas  |
| 593 | 1 | EUTC_PSEAE | Q9hx02 | pseudomonas   | 666 | 6 | 0.8 | 303 | 1 | PYRD_PYRAB | Q940Y6 | pyrococcus   |
| 594 | 1 | FLJL_CAUCR | P18914 | caulobacter   | 667 | 6 | 0.8 | 303 | 1 | PYRD_PYRHO | O59185 | pyrococcus   |
| 595 | 1 | YD66_MYCTU | Q11036 | mycobacteri   | 668 | 6 | 0.8 | 303 | 1 | V112_FOWPV | Q9J523 | fowlpox vir  |
| 596 | 1 | DAPD_ACTPL | P41396 | actinobacill  | 669 | 6 | 0.8 | 304 | 1 | HEM3_AQUAE | O66621 | aquifex aeo  |
| 597 | 1 | DAPD_EUCAI | P57323 | buchnera ap   | 670 | 6 | 0.8 | 304 | 1 | HEM3_XANAC | Q8ppr3 | xanthomonas  |
| 598 | 1 | DAPD_ECOLI | P03948 | escherichia   | 671 | 6 | 0.8 | 304 | 1 | HEM3_XANCP | Q8p536 | xanthomonas  |
| 599 | 1 | DAPD_MYCOB | P56220 | mycobacteri   | 672 | 6 | 0.8 | 304 | 1 | Y007_MYCTU | P71575 | mycobacteri  |
| 600 | 1 | GLO2_YEAST | Q05584 | saccharomyc   | 673 | 6 | 0.8 | 305 | 1 | HEM3_XYLPA | Q9pcx7 | xyliella fas |
| 601 | 1 | MEPA_ECOLI | P14007 | escherichia   | 674 | 6 | 0.8 | 305 | 1 | Y452_BUCAL | P57527 | buchnera ap  |
| 602 | 1 | MURI_OCEIH | Q8ep14 | oceanobacill  | 675 | 6 | 0.8 | 305 | 1 | YBBK_ECOLI | P77367 | escherichia  |
| 603 | 1 | T2X1_XANCR | P96189 | xanthomonas   | 676 | 6 | 0.8 | 306 | 1 | HEM3_HELPJ | Q9xmj7 | helicobacte  |
| 604 | 1 | DAPD_HABIN | P45284 | haemophilus   | 677 | 6 | 0.8 | 306 | 1 | HEM3_HELPJ | P56140 | helicobacte  |
| 605 | 1 | ACT_CALFI  | Q92192 | calanus fin   | 678 | 6 | 0.8 | 307 | 1 | REP_BACSP  | P36229 | campylobact  |
| 606 | 1 | DAP2_ANASP | Q8yvd0 | anabaena sp   | 679 | 6 | 0.8 | 307 | 1 | VER3_YEAST | P36229 | bacillus sp  |
| 607 | 1 | YIFB_BACSU | P37497 | bacillus su   | 680 | 6 | 0.8 | 307 | 1 | ACPI_ENTHI | P40056 | saccharomyc  |
| 608 | 1 | ANFD_HELGE | O68955 | helicobacter  | 681 | 6 | 0.8 | 308 | 1 | HEM3_STAAU | P36184 | entamoeba h  |
| 609 | 1 | SQ22_CANAL | P87218 | candida alb   | 682 | 6 | 0.8 | 308 | 1 | HEM3_STAAU | Q99tj1 | staphylococ  |
| 610 | 1 | V236_FOWPV | P14365 | fowlpox vir   | 683 | 6 | 0.8 | 308 | 1 | HEM3_STAAU | Q34090 | staphylococ  |
| 611 | 1 | VG85_BPPI2 | P20344 | bacterioph    | 684 | 6 | 0.8 | 308 | 1 | HEM3_STAAU | Q8rw74 | staphylococ  |
| 612 | 1 | VG85_BPPI2 | P07532 | bacterioph    | 685 | 6 | 0.8 | 308 | 1 | PQOB_KLEFN | P27504 | klebsiella   |
| 613 | 1 | APAH_PSEAE | Q915u7 | pseudomonas   | 686 | 6 | 0.8 | 309 | 1 | ACT3_ECHGR | Q03342 | echinococcu  |
| 614 | 1 | HISI_METKA | Q8tyd5 | methanopyru   | 687 | 6 | 0.8 | 309 | 1 | CYP9_CABEL | O09637 | caenorhabdi  |
| 615 | 1 | CDX4_HUMAN | O14627 | homo sapien   | 688 | 6 | 0.8 | 309 | 1 | NUSG_STRGB | P52852 | streptomyce  |
| 616 | 1 | CELA_ACEXY | P27897 | acetobacter   | 689 | 6 | 0.8 | 309 | 1 | RLA0_BOVIN | Q95140 | bos taurus   |
| 617 | 1 | FOCA_HABIN | P43756 | haemophilus   | 690 | 6 | 0.8 | 309 | 1 | UCP2_FIG   | Q97562 | sus scrofa   |

|     |   |     |     |   |            |        |              |     |   |     |     |   |            |        |              |
|-----|---|-----|-----|---|------------|--------|--------------|-----|---|-----|-----|---|------------|--------|--------------|
| 691 | 6 | 0.8 | 309 | 1 | URIC_ARATH | O04420 | arabidopsis  | 764 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P32547 | porcine rot  |
| 692 | 6 | 0.8 | 309 | 1 | YBGR_HAHN  | P44298 | haemophilus  | 765 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P32548 | porcine rot  |
| 693 | 6 | 0.8 | 309 | 1 | YBGR_ECOLI | P76369 | escherichia  | 766 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P27423 | porcine rot  |
| 694 | 6 | 0.8 | 310 | 1 | CCW7_YEAST | P47000 | saccharomyc  | 767 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P32549 | porcine rot  |
| 695 | 6 | 0.8 | 310 | 1 | SIAL_BOVIN | Q28862 | bos taurus   | 768 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P17466 | porcine rot  |
| 696 | 6 | 0.8 | 310 | 1 | YN00_MYCTU | Q50665 | mycobacteri  | 769 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P12476 | phesus rota  |
| 697 | 6 | 0.8 | 311 | 1 | PABD_BUCBP | Q89ah0 | buchnera ap  | 770 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P03533 | smian l1 r   |
| 698 | 6 | 0.8 | 311 | 1 | HMS3_CHLVI | P28464 | cholorobus   | 771 | 6 | 0.8 | 326 | 1 | YR75_ECOLI | P12314 | escherichia  |
| 699 | 6 | 0.8 | 311 | 1 | KHSE_SULSO | Q97w70 | sulfolobus   | 772 | 6 | 0.8 | 327 | 1 | ACT3_ARTSK | P18602 | artemia sp.  |
| 700 | 6 | 0.8 | 312 | 1 | COLV_YERPE | P17811 | yersinia pe  | 773 | 6 | 0.8 | 327 | 1 | MTN1_HAEHA | P05102 | haemophilus  |
| 701 | 6 | 0.8 | 312 | 1 | HMS3_CHUTE | Q8kc14 | cholorobium  | 774 | 6 | 0.8 | 327 | 1 | SVK_THETN  | Q8-9x8 | thermoanaer  |
| 702 | 6 | 0.8 | 312 | 1 | RIR2_RSIV  | Q9qt12 | red sea bre  | 775 | 6 | 0.8 | 327 | 1 | YIIZ_SALTY | P43020 | salmonella   |
| 703 | 6 | 0.8 | 313 | 1 | TFS2_DROME | P20232 | drosophila   | 776 | 6 | 0.8 | 328 | 1 | HMD1_HUMAN | Q99j22 | homo sapien  |
| 704 | 6 | 0.8 | 314 | 1 | HCDD_PIG   | P00348 | sus scrofa   | 777 | 6 | 0.8 | 329 | 1 | PEI7_ARATH | Q99j22 | arabidopsis  |
| 705 | 6 | 0.8 | 314 | 1 | LIPA_MYCLE | Q32962 | mycobacteri  | 778 | 6 | 0.8 | 329 | 1 | YIID_ECOLI | P32148 | escherichia  |
| 706 | 6 | 0.8 | 314 | 1 | PYRB_DEIRA | Q9rvco | deinococcus  | 779 | 6 | 0.8 | 330 | 1 | PE70_ARATH | Q9fmi7 | arabidopsis  |
| 707 | 6 | 0.8 | 314 | 1 | SINA_DROME | P21461 | drosophila   | 780 | 6 | 0.8 | 331 | 1 | TAL_HUMAN  | P17542 | homo sapien  |
| 708 | 6 | 0.8 | 314 | 1 | SINA_DROVI | P23304 | drosophila   | 781 | 6 | 0.8 | 331 | 1 | TXJ6_CABEL | Q23622 | caenorhabdi  |
| 709 | 6 | 0.8 | 315 | 1 | MRAW_BUCBP | P59522 | buchnera ap  | 782 | 6 | 0.8 | 333 | 1 | GRP8_HUMAN | P48146 | homo sapien  |
| 710 | 6 | 0.8 | 315 | 1 | RUVB_MYCFU | Q98pr1 | mycoplasma   | 783 | 6 | 0.8 | 333 | 1 | MRAW_ZYMOO | Q9reg9 | zymomonas m  |
| 711 | 6 | 0.8 | 315 | 1 | TGFI_BOVIN | P18341 | bos taurus   | 784 | 6 | 0.8 | 333 | 1 | PHL2_BACCE | P11889 | bacillus ce  |
| 712 | 6 | 0.8 | 315 | 1 | RLAO_CHICK | P47826 | gallus gall  | 785 | 6 | 0.8 | 333 | 1 | PHL3_BACCE | P33377 | bacillus ce  |
| 713 | 6 | 0.8 | 316 | 1 | RSP4_TRIGR | P38980 | tripeustes   | 786 | 6 | 0.8 | 333 | 1 | YACF_BACSU | P37567 | bacillus su  |
| 714 | 6 | 0.8 | 316 | 1 | TAL_HELPV  | Q2jcs  | helicobacte  | 787 | 6 | 0.8 | 334 | 1 | HEM3_RALSO | Q8xwv3 | ralstonia s  |
| 715 | 6 | 0.8 | 316 | 1 | TAL_HELPV  | P56108 | helicobacte  | 788 | 6 | 0.8 | 335 | 1 | PAPG_ECOLI | P13720 | escherichia  |
| 716 | 6 | 0.8 | 316 | 1 | Y007_CHLTR | Q84010 | chlamydia c  | 789 | 6 | 0.8 | 335 | 1 | PSG5_HUMAN | Q15238 | homo sapien  |
| 717 | 6 | 0.8 | 317 | 1 | CBXP_ALCEU | Q04540 | alcaligenes  | 790 | 6 | 0.8 | 336 | 1 | ACT2_SOLTU | P93374 | nicotiana t  |
| 718 | 6 | 0.8 | 317 | 1 | MRAY_CLOAB | Q97h86 | clostridium  | 791 | 6 | 0.8 | 336 | 1 | ACT2_TOBAC | P93375 | nicotiana t  |
| 719 | 6 | 0.8 | 317 | 1 | RLAO_HUMAN | P05388 | homo sapien  | 792 | 6 | 0.8 | 337 | 1 | APL2_HUMAN | Q9bqes | homo sapien  |
| 720 | 6 | 0.8 | 317 | 1 | RLAO_MOUSE | P14869 | mus musculus | 793 | 6 | 0.8 | 337 | 1 | SAT3_YEAST | P06844 | saccharomyc  |
| 721 | 6 | 0.8 | 317 | 1 | RLAO_RAT   | P19945 | rattus norv  | 794 | 6 | 0.8 | 337 | 1 | TAT_HTLV2  | P03410 | human t-cel  |
| 722 | 6 | 0.8 | 317 | 1 | TAL_YERPE  | Q8zin2 | yersinia pe  | 795 | 6 | 0.8 | 337 | 1 | YBJS_ECOLI | P75821 | escherichia  |
| 723 | 6 | 0.8 | 318 | 1 | DHP2_MYCTU | Q53308 | mycobacteri  | 796 | 6 | 0.8 | 337 | 1 | YF02_CLOPE | P58698 | clostridium  |
| 724 | 6 | 0.8 | 318 | 1 | KDGT_XANAC | Q8pdq6 | xanthomonas  | 797 | 6 | 0.8 | 338 | 1 | MTBA_METAC | P58699 | methanosarc  |
| 725 | 6 | 0.8 | 318 | 1 | VAIL_VACCC | P20988 | vaccinia vi  | 798 | 6 | 0.8 | 338 | 1 | MTBA_METBA | Q30640 | methanosarc  |
| 726 | 6 | 0.8 | 318 | 1 | Y340_METJA | Q58350 | methanococc  | 799 | 6 | 0.8 | 338 | 1 | MTBA_METBA | P58984 | methanosarc  |
| 727 | 6 | 0.8 | 319 | 1 | MOCB_SYNP7 | Q56208 | synechococc  | 800 | 6 | 0.8 | 338 | 1 | Y941_METJA | Q57711 | methanococc  |
| 728 | 6 | 0.8 | 319 | 1 | TRXB_BUCAP | P57399 | buchnera ap  | 801 | 6 | 0.8 | 338 | 1 | COND_NEUTR | P19463 | neurospora   |
| 729 | 6 | 0.8 | 319 | 1 | TRXB_BUCAP | P81433 | buchnera ap  | 802 | 6 | 0.8 | 340 | 1 | KHVB_ECOLI | P00557 | escherichia  |
| 730 | 6 | 0.8 | 319 | 1 | VAIL_VARV  | P3836  | variola vir  | 803 | 6 | 0.8 | 341 | 1 | KHVB_ECOLI | P14341 | varicella-z  |
| 731 | 6 | 0.8 | 319 | 1 | YER1_SCHPO | O14084 | schizosacch  | 804 | 6 | 0.8 | 341 | 1 | KITH_VZV4  | P14342 | varicella-z  |
| 732 | 6 | 0.8 | 320 | 1 | FE22_HUMAN | Q9uhy8 | homo sapien  | 805 | 6 | 0.8 | 341 | 1 | KITH_VZV7  | P09250 | varicella-z  |
| 733 | 6 | 0.8 | 322 | 1 | ACT_PROCL  | P45521 | procambur    | 806 | 6 | 0.8 | 341 | 1 | KITH_VZV7  | P14343 | varicella-z  |
| 734 | 6 | 0.8 | 322 | 1 | ADT_SCHPO  | Q9188  | schizosacch  | 807 | 6 | 0.8 | 341 | 1 | KITH_VZV7  | P14344 | varicella-z  |
| 735 | 6 | 0.8 | 322 | 1 | RLAO_LEICH | P39096 | leishmania   | 808 | 6 | 0.8 | 341 | 1 | KITH_VZVW  | P55771 | homo sapien  |
| 736 | 6 | 0.8 | 322 | 1 | SLP1_DROME | P32030 | drosophila   | 809 | 6 | 0.8 | 341 | 1 | PAX9_HUMAN | Q03178 | saccharomyc  |
| 737 | 6 | 0.8 | 323 | 1 | KITH_HSVBH | P22643 | bovine herp  | 810 | 6 | 0.8 | 341 | 1 | PIR1_YEAST | Q9x7f4 | methylobact  |
| 738 | 6 | 0.8 | 323 | 1 | RLAO_LEIN  | P32097 | leishmania   | 811 | 6 | 0.8 | 342 | 1 | COBT_YETSO | Q9x7f4 | methylobact  |
| 739 | 6 | 0.8 | 324 | 1 | FE22_RAT   | P97578 | rattus norv  | 812 | 6 | 0.8 | 342 | 1 | FLGI_HELPV | Q25028 | helicobacte  |
| 740 | 6 | 0.8 | 324 | 1 | Y734_CHLPN | Q9z7h1 | chlamydia p  | 813 | 6 | 0.8 | 342 | 1 | FLGI_HELPV | C25028 | helicobacte  |
| 741 | 6 | 0.8 | 325 | 1 | MRAY_BACHD | Q9k9s6 | bacillus ha  | 814 | 6 | 0.8 | 342 | 1 | PAX9_MOUSE | P47242 | mus musculus |
| 742 | 6 | 0.8 | 325 | 1 | Y132_CHLPN | Q9z949 | chlamydia p  | 815 | 6 | 0.8 | 343 | 1 | DDL_CLOAB  | Q97f58 | clostridium  |
| 743 | 6 | 0.8 | 326 | 1 | ATP3_IPOBA | P26360 | ipomoea bat  | 816 | 6 | 0.8 | 343 | 1 | Y535_METJA | Q57955 | methanococc  |
| 744 | 6 | 0.8 | 326 | 1 | HXA1_HETFR | Q9ia19 | heterodontu  | 817 | 6 | 0.8 | 343 | 1 | YMD4_CABEL | P34458 | caenorhabdi  |
| 745 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | Q01252 | bovine rota  | 818 | 6 | 0.8 | 344 | 1 | MALR_STRCO | P72396 | streptomyce  |
| 746 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P17968 | bovine rota  | 819 | 6 | 0.8 | 344 | 1 | NUZM_RANCA | P72396 | streptomyce  |
| 747 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | Q00253 | bovine rota  | 820 | 6 | 0.8 | 345 | 1 | NUZM_XENLA | P16673 | rana catesb  |
| 748 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P17700 | bovine rota  | 821 | 6 | 0.8 | 345 | 1 | REG1_STRLI | P03894 | xenopus lae  |
| 749 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P30210 | bovine rota  | 822 | 6 | 0.8 | 345 | 1 | YHDE_ECOLI | P72469 | streptomyce  |
| 750 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P25176 | equine rota  | 823 | 6 | 0.8 | 345 | 1 | YHDE_ECOLI | P46853 | escherichia  |
| 751 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | Q9ipd4 | equine rota  | 824 | 6 | 0.8 | 346 | 1 | LICH_ACILW | P77585 | escherichia  |
| 752 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | Q9187  | human rotav  | 825 | 6 | 0.8 | 346 | 1 | NUZM_ANAAC | Q9x284 | acinetobact  |
| 753 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P17071 | human rotav  | 826 | 6 | 0.8 | 346 | 1 | NUZM_ANAAC | O63767 | anas acuta   |
| 754 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P1285  | human rotav  | 827 | 6 | 0.8 | 346 | 1 | NUZM_ANAAC | O63794 | anas ameri   |
| 755 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P11854 | human rotav  | 828 | 6 | 0.8 | 346 | 1 | NUZM_ANAAC | O63797 | anas faicat  |
| 756 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | Q08779 | human rotav  | 829 | 6 | 0.8 | 346 | 1 | NUZM_ANAAC | O63798 | anas formos  |
| 757 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P11855 | human rotav  | 830 | 6 | 0.8 | 346 | 1 | NUZM_ANAAC | O63798 | anas formos  |
| 758 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P10501 | human rotav  | 831 | 6 | 0.8 | 346 | 1 | NUZM_STRCA | Q21398 | struthio ca  |
| 759 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P11856 | human rotav  | 832 | 6 | 0.8 | 346 | 1 | PPF_DICTH  | Q9kh71 | d pyrophosp  |
| 760 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P12545 | porcine rot  | 833 | 6 | 0.8 | 346 | 1 | QUEA_NEIMA | Q9jw05 | neisseria m  |
| 761 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P32546 | porcine rot  | 834 | 6 | 0.8 | 346 | 1 | QUEA_NEIMA | Q9jw05 | neisseria m  |
| 762 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P09365 | porcine rot  | 835 | 6 | 0.8 | 346 | 1 | RECA_THIFE | P16238 | thiobacillu  |
| 763 | 6 | 0.8 | 326 | 1 | VS09_ROTPE | P08406 | porcine rot  | 836 | 6 | 0.8 | 346 | 1 | TBCC_HUMAN | Q15614 | homo sapien  |

|     |   |            |     |     |   |                     |     |   |            |     |     |   |                    |
|-----|---|------------|-----|-----|---|---------------------|-----|---|------------|-----|-----|---|--------------------|
| 837 | 1 | TH13_SCHPO | 346 | 0.8 | 6 | P36597 schizosacch  | 910 | 1 | CARA_METWA | 368 | 0.8 | 6 | Q8Q0u4 methanosarc |
| 838 | 1 | WRS4_ARATH | 346 | 0.8 | 6 | Q93wu8 arabidopsis  | 911 | 1 | GMD1_CABEL | 368 | 0.8 | 6 | Q18601 caenorhabdi |
| 839 | 1 | HTPX_PRAE  | 347 | 0.8 | 6 | Q82c30 pyrococcus   | 912 | 1 | MILD_STAAM | 369 | 0.8 | 6 | Q998a1 staphylococ |
| 840 | 1 | PRIM_PRAE  | 347 | 0.8 | 6 | Q9pnh1 pyrococcus   | 913 | 1 | Y264_SYNY3 | 369 | 0.8 | 6 | P73879 synchocyst  |
| 841 | 1 | FLGI_CAMJE | 348 | 0.8 | 6 | Q9pjh8 campylobact  | 914 | 1 | ACT5_DIPDE | 371 | 0.8 | 6 | P53458 diphyllobot |
| 842 | 1 | PEPQ_PPRFU | 348 | 0.8 | 6 | P81535 pyrococcus   | 915 | 1 | FLGI_RHOSH | 371 | 0.8 | 6 | P58204 rhodobacter |
| 843 | 1 | YF91_SYNY3 | 348 | 0.8 | 6 | P74598 synchocyst   | 916 | 1 | PGLR_PENVA | 371 | 0.8 | 6 | O42824 penicillium |
| 844 | 1 | YR35_STRFR | 348 | 0.8 | 6 | P20186 streptomyce  | 917 | 1 | RV02_YEAST | 371 | 0.8 | 6 | P12687 saccharomyc |
| 845 | 1 | MOAA_CAUCR | 349 | 0.8 | 6 | Q9ac48 caulobacter  | 918 | 1 | DBPA_HUMAN | 372 | 0.8 | 6 | P16989 homo sapien |
| 846 | 1 | OPSP_COLL  | 349 | 0.8 | 6 | P51476 columba liv  | 919 | 1 | DDL_MYCTM  | 373 | 0.8 | 6 | Q929n0 mycobacteri |
| 847 | 1 | RECA_BACHD | 349 | 0.8 | 6 | Q9kaa7 bacillus ha  | 920 | 1 | DDL_MYCTM  | 373 | 0.8 | 6 | P55114 mycobacteri |
| 848 | 1 | RS6_AEDAL  | 349 | 0.8 | 6 | Q9u762 aedes albop  | 921 | 1 | HMZ1_DROSU | 373 | 0.8 | 6 | Q24648 drosophila  |
| 849 | 1 | SGF1_BOMMO | 349 | 0.8 | 6 | Q17411 bombyx mori  | 922 | 1 | MGEL_HUMAN | 373 | 0.8 | 6 | Q9ubf1 homo sapien |
| 850 | 1 | YGB0_ECOLI | 349 | 0.8 | 6 | O57261 escherichia  | 923 | 1 | TGF1_CHICK | 373 | 0.8 | 6 | P09531 gallus gall |
| 851 | 1 | CMTG_PSEPU | 350 | 0.8 | 6 | O51983 pseudomonas  | 924 | 1 | CHEB_PSEPU | 374 | 0.8 | 6 | O52262 pseudomonas |
| 852 | 1 | PLEK_HUMAN | 350 | 0.8 | 6 | P08567 homo sapien  | 925 | 1 | FLAE_VIBRA | 374 | 0.8 | 6 | Q24ba2 vibrio para |
| 853 | 1 | PLEK_MOUSE | 350 | 0.8 | 6 | Q9j1k5 mus musculu  | 926 | 1 | PROB_STRCO | 375 | 0.8 | 6 | Q9rdj9 streptomyce |
| 854 | 1 | XPFP_HUMAN | 350 | 0.8 | 6 | O75895 homo sapien  | 927 | 1 | ACT1_ACACA | 375 | 0.8 | 6 | P02578 acanthamoeb |
| 855 | 1 | XYLA_PSEPU | 350 | 0.8 | 6 | P21394 pseudomonas  | 928 | 1 | ACT1_DICDI | 375 | 0.8 | 6 | P02577 dictyosteli |
| 856 | 1 | COAT_TCV   | 351 | 0.8 | 6 | P06663 turnip crin  | 929 | 1 | ACT1_FUGRU | 375 | 0.8 | 6 | P53484 figu rubrip |
| 857 | 1 | HEM3_RICCN | 351 | 0.8 | 6 | Q22hr5 rickettsia   | 930 | 1 | ACT1_PNECA | 375 | 0.8 | 6 | P43239 pneumocysti |
| 858 | 1 | HOMN_ALCEU | 351 | 0.8 | 6 | P23516 alcaligenes  | 931 | 1 | ACT1_SCHCO | 375 | 0.8 | 6 | Q9y702 schizophyll |
| 859 | 1 | MTD5_DACSA | 351 | 0.8 | 6 | O50185 dactylococc  | 932 | 1 | ACT1_SUIBO | 375 | 0.8 | 6 | Q9y701 suillus bov |
| 860 | 1 | PEPQ_PVRHO | 351 | 0.8 | 6 | O58885 pyrococcus   | 933 | 1 | ACT2_FUGRU | 375 | 0.8 | 6 | P53485 figu rubrip |
| 861 | 1 | PEXE_PICAN | 351 | 0.8 | 6 | P78223 pichia angu  | 934 | 1 | ACT2_SCHCO | 375 | 0.8 | 6 | Q9y896 schizophyll |
| 862 | 1 | VP39_NPVOP | 352 | 0.8 | 6 | P17500 oryza pseu   | 935 | 1 | ACT2_SUIBO | 375 | 0.8 | 6 | Q9y707 suillus bov |
| 863 | 1 | HEMZ_PROFR | 352 | 0.8 | 6 | P72183 propionibac  | 936 | 1 | ACT3_FUGRU | 375 | 0.8 | 6 | P53486 figu rubrip |
| 864 | 1 | RECA_ACTAC | 352 | 0.8 | 6 | Q9j1p9 actinobacil  | 937 | 1 | ACTA_PHYPO | 375 | 0.8 | 6 | P02576 physarum po |
| 865 | 1 | RLAO_HALCU | 352 | 0.8 | 6 | P17006 halobacteri  | 938 | 1 | ACTB_CRIGR | 375 | 0.8 | 6 | P48975 cricetus    |
| 866 | 1 | RLAO_HALN1 | 352 | 0.8 | 6 | P13553 halobacteri  | 939 | 1 | ACTB_CYPCA | 375 | 0.8 | 6 | P22714 cyprinus ca |
| 867 | 1 | SUB1_SYNY3 | 352 | 0.8 | 6 | Q01303 synchocyst   | 940 | 1 | ACTB_HUMAN | 375 | 0.8 | 6 | P02570 homo sapien |
| 868 | 1 | WN41_BRARE | 352 | 0.8 | 6 | P47793 brachydanio  | 941 | 1 | ACTB_ORYLA | 375 | 0.8 | 6 | P79818 oryztas lat |
| 869 | 1 | VP10_RDVA  | 353 | 0.8 | 6 | Q85447 rice dwarf   | 942 | 1 | ACTB_RABIT | 375 | 0.8 | 6 | P29751 oryctolagus |
| 870 | 1 | MORG_BUCAP | 353 | 0.8 | 6 | Q8k9t4 buchnera ap  | 943 | 1 | ACTB_SALSA | 375 | 0.8 | 6 | O42161 salmo salar |
| 871 | 1 | PON2_MOUSE | 354 | 0.8 | 6 | Q62086 mus musculu  | 944 | 1 | ACTC_BRABE | 375 | 0.8 | 6 | Q93129 branchiosto |
| 872 | 1 | ALR_RICCN  | 355 | 0.8 | 6 | Q92jd9 rickettsia   | 945 | 1 | ACTC_BRABE | 375 | 0.8 | 6 | Q93131 branchiosto |
| 873 | 1 | LPXD_AGR5  | 355 | 0.8 | 6 | Q8uf15 agrobacteri  | 946 | 1 | ACTC_BRABL | 375 | 0.8 | 6 | O91131 branchiosto |
| 874 | 1 | RECA_SYNEL | 355 | 0.8 | 6 | O8d170 synchococc   | 947 | 1 | ACTC_BRALA | 375 | 0.8 | 6 | O17503 branchiosto |
| 875 | 1 | GLPO_TREPA | 356 | 0.8 | 6 | O30405 treponema p  | 948 | 1 | ACTG_CEPAC | 375 | 0.8 | 6 | Q9uuv9 cephalospor |
| 876 | 1 | TRM1_AQUAE | 356 | 0.8 | 6 | O67274 aquifex aeo  | 949 | 1 | ACTG_EMENT | 375 | 0.8 | 6 | P20359 emeritella  |
| 877 | 1 | FIBP_CRAAE | 357 | 0.8 | 6 | Q9j1l9 mus musculu  | 950 | 1 | ACTG_HUMAN | 375 | 0.8 | 6 | P02571 homo sapien |
| 878 | 1 | FIBP_MOUSE | 357 | 0.8 | 6 | Q9j1l9 mus musculu  | 951 | 1 | ACTG_PENCH | 375 | 0.8 | 6 | Q9urs0 penicillium |
| 879 | 1 | LIBP_MYXXA | 357 | 0.8 | 6 | Q9x6x4 myxococcus   | 952 | 1 | ACT_BOTCI  | 375 | 0.8 | 6 | O13419 botrytis ci |
| 880 | 1 | RF1_CHLPN  | 357 | 0.8 | 6 | Q92968 chlamydia p  | 953 | 1 | ACT_COPCI  | 375 | 0.8 | 6 | Q9uvx4 coprinus ci |
| 881 | 1 | CKR3_CAVPO | 358 | 0.8 | 6 | Q92213 cavia porce  | 954 | 1 | ACT_KLULA  | 375 | 0.8 | 6 | P17128 kluyveromyc |
| 882 | 1 | SYFA_STRCO | 358 | 0.8 | 6 | Q92213 cavia porce  | 955 | 1 | ACT_NEUCR  | 375 | 0.8 | 6 | P78711 neurospora  |
| 883 | 1 | ADD1_STRCO | 359 | 0.8 | 6 | O80055 streptomyce  | 956 | 1 | ACT_PHARH  | 375 | 0.8 | 6 | P53689 phaffia rho |
| 884 | 1 | KLFS_HUMAN | 359 | 0.8 | 6 | Q9x7c2 streptomyce  | 957 | 1 | ACT_SCHPO  | 375 | 0.8 | 6 | P10989 schizosacch |
| 885 | 1 | RF1_CHLMU  | 360 | 0.8 | 6 | O95600 homo sapien  | 958 | 1 | ACT_YEAST  | 375 | 0.8 | 6 | Q9uvf3 yarrowia li |
| 886 | 1 | AROB_THEVO | 360 | 0.8 | 6 | Q9p116 chlamydia m  | 959 | 1 | ACT1_AEDAE | 376 | 0.8 | 6 | P49128 aedes aegyp |
| 887 | 1 | IHA_BOVIN  | 360 | 0.8 | 6 | Q978s6 thermoplasma | 960 | 1 | ACT1_AEDAE | 376 | 0.8 | 6 | P02579 saccharomyc |
| 888 | 1 | HEM3_MOUSE | 361 | 0.8 | 6 | P07994 bos taurus   | 961 | 1 | ACT1_AEDAE | 376 | 0.8 | 6 | P18600 artemia sp  |
| 889 | 1 | HEM3_RAT   | 361 | 0.8 | 6 | P22907 mus musculu  | 962 | 1 | ACT1_BOMMO | 376 | 0.8 | 6 | P07836 bombyx mori |
| 890 | 1 | IFP3_MOUSE | 361 | 0.8 | 6 | Q9dch4 mus musculu  | 963 | 1 | ACT1_CABEL | 376 | 0.8 | 6 | P10983 caenorhabdi |
| 891 | 1 | FENP_ORYSA | 362 | 0.8 | 6 | P41344 oryza sativ  | 964 | 1 | ACT1_DROME | 376 | 0.8 | 6 | P10987 drosophila  |
| 892 | 1 | MSPI_YEAST | 362 | 0.8 | 6 | P28737 saccharomyc  | 965 | 1 | ACT1_HELER | 376 | 0.8 | 6 | P53462 heliocidari |
| 893 | 1 | CYSP_PEA   | 363 | 0.8 | 6 | P25804 pisum sativ  | 966 | 1 | ACT1_LYTP1 | 376 | 0.8 | 6 | P53465 lytechinus  |
| 894 | 1 | AAT_PYROK  | 364 | 0.8 | 6 | Q93744 pyrococcus   | 967 | 1 | ACT1_ONCVO | 376 | 0.8 | 6 | P30162 onchocerca  |
| 895 | 1 | FIBP_HUMAN | 364 | 0.8 | 6 | P23610 homo sapien  | 968 | 1 | ACT1_PODCA | 376 | 0.8 | 6 | P41112 podocoryne  |
| 896 | 1 | LEU3_RICUE | 364 | 0.8 | 6 | Q43427 homo sapien  | 969 | 1 | ACT1_SACKO | 376 | 0.8 | 6 | O84999 sacoglossu  |
| 897 | 1 | PAR2_HUMAN | 364 | 0.8 | 6 | Q9evh5 buchnera ap  | 970 | 1 | ACT1_SCHMA | 376 | 0.8 | 6 | P53470 schistosoma |
| 898 | 1 | CATD_SHEEP | 365 | 0.8 | 6 | Q9nb11 homo sapien  | 971 | 1 | ACT1_STRPN | 376 | 0.8 | 6 | P10990 strongyloce |
| 899 | 1 | HEMA_ECOLI | 365 | 0.8 | 6 | Q9mzsh ovis aries   | 972 | 1 | ACT2_ARTSX | 376 | 0.8 | 6 | P18601 artemia sp  |
| 900 | 1 | RECA_SPIPL | 365 | 0.8 | 6 | P77258 escherichia  | 973 | 1 | ACT2_BACDO | 376 | 0.8 | 6 | P45885 bactroceca  |
| 901 | 1 | SOXE_RHOSH | 365 | 0.8 | 6 | P48293 spirulina p  | 974 | 1 | ACT2_BOMMO | 376 | 0.8 | 6 | P07837 bombyx mori |
| 902 | 1 | YKQ7_CABEL | 365 | 0.8 | 6 | P54997 rhodococcus  | 975 | 1 | ACT2_CASEL | 376 | 0.8 | 6 | P10984 caenorhabdi |
| 903 | 1 | ACTD_PHYPO | 366 | 0.8 | 6 | P34302 caenorhabdi  | 976 | 1 | ACT2_DIPDE | 376 | 0.8 | 6 | P53456 diphyllobot |
| 904 | 1 | IHA_MOUSE  | 366 | 0.8 | 6 | P24263 physarum po  | 977 | 1 | ACT2_DROME | 376 | 0.8 | 6 | P02572 drosophila  |
| 905 | 1 | MRAY_COREP | 366 | 0.8 | 6 | Q04997 mus musculu  | 978 | 1 | ACT2_ECHGR | 376 | 0.8 | 6 | Q03341 echinococcu |
| 906 | 1 | MRAY_COREP | 366 | 0.8 | 6 | Q8fnt7 corynebacte  | 979 | 1 | ACT2_LUNTE | 376 | 0.8 | 6 | P92176 lumbricus t |
| 907 | 1 | PGLR_PENDI | 367 | 0.8 | 6 | Q8nnn2 corynebacte  | 980 | 1 | ACT2_LYTP1 | 376 | 0.8 | 6 | P53466 lytechinus  |
| 908 | 1 | CARA_METAC | 368 | 0.8 | 6 | Q8trny3 methanosarc | 981 | 1 | ACT2_ONCVO | 376 | 0.8 | 6 | P30163 onchocerca  |
| 909 | 1 |            |     |     |   |                     | 982 | 1 | ACT2_SACKO | 376 | 0.8 | 6 | O18500 sacoglossu  |
|     |   |            |     |     |   |                     |     |   | ACT2_SCHMA |     |     |   | P53471 schistosoma |

983 6 0.8 376 1 ACT3\_BACDO P45886 bactrocera  
984 6 0.8 376 1 ACT3\_BOMVO P04829 bombyx mori  
985 6 0.8 376 1 ACT3\_DROME P53501 drosophila  
986 6 0.8 376 1 ACT3\_HELAM Q25010 helicoverpa  
987 6 0.8 376 1 ACT3\_PODCA P41113 podocoryne  
988 6 0.8 376 1 ACT3\_ARTSX P18603 artemia sp.  
989 6 0.8 376 1 ACT4\_BOMVO Q27250 bombyx mori  
990 6 0.8 376 1 ACT4\_CAEEL P10986 caenorhabdi  
991 6 0.8 376 1 ACT4\_DROME P02574 drosophila  
992 6 0.8 376 1 ACT5\_BACDO P45887 bactrocera  
993 6 0.8 376 1 ACT5\_CHICK P53478 gallus gall  
994 6 0.8 376 1 ACT5\_DROME P10981 drosophila  
995 6 0.8 376 1 ACT5\_XENLA P53505 xenopus lae  
996 6 0.8 376 1 ACT6\_DROME P02575 drosophila  
997 6 0.8 376 1 ACT6\_XENLA P53506 xenopus lae  
998 6 0.8 376 1 ACTA\_STRPU P53472 strongyloce  
999 6 0.8 376 1 ACTB\_STRPU P53473 strongyloce  
1000 6 0.8 376 1 ACTB\_XENBO P53475 xenopus bor

ALIGNMENTS

RESULT 1  
OMC NEIGO  
ID OMC NEIGO STANDARD; PRT; 711 AA.  
AC P35819;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE Outer membrane protein OMC precursor.  
GN OMC.  
OS Neisseria gonorrhoeae.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
OC Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=485;  
RN [1]\_TaxID=485;  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2686;  
RA Tsai W.M., Larsen S.H., Wilde C.E.;  
RT "Cloning and DNA sequence of the omc gene encoding the outer membrane protein-macromolecular complex from Neisseria gonorrhoeae.";  
RL Infect. Immun. 57:2653-2659(1989).  
CC -!- SUBUNIT: FORMS A HIGH MACROMOLECULAR COMPLEX IN THE OUTER MEMBRANE.  
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE, ASSOCIATED TO THE MEMBRANE THROUGH ITS C-TERMINAL.  
CC -!- SIMILARITY: BELONGS TO THE PULD/OUTD/EXED/XPSD FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: L19944; AAA25456.1;  
DR InterPro: IPR004846; GSPII/IIIprotein.  
DR InterPro: IPR004845; GSPIIproteinC.  
DR Pfam: PF00263; GSPII\_III; 1.  
DR Pfam: PF03958; GSPII\_III N; 1.  
DR PROSITE: PS00875; T2SP D; FALSE NEG.  
KW Transport; Outer membrane; Signal.  
FT SIGNAL 1 24 POTENTIAL.  
FT CHAIN 25 711 OUTER MEMBRANE PROTEIN OMC.  
SQ SEQUENCE 711 AA; 77483 MW; 249CF9D8B65F9F0 CRC64;

Query Match 18.5%; Score 142; DB 1; Length 711;  
Best Local Similarity 100.0%; Pred. No. 2.7e-135;  
Matches 142; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNTKLTIIISGLFVATAAQTASAGNITDIKVSSLPNKQIKVKSFDKEIVNPTGFTVSS 60  
DB 1 MNTKLTIIISGLFVATAAQTASAGNITDIKVSSLPNKQIKVKSFDKEIVNPTGFTVSS 60  
QY 61 PARIALDFEOTGISMDOQVLEAYADPULSKISAQNSSRARLVNLNKPQCYNTEVRGNKV 120  
DB 61 PARIALDFEOTGISMDOQVLEAYADPULSKISAQNSSRARLVNLNKPQCYNTEVRGNKV 120  
QY 121 WIFINESDDTVSAPARPAVAAA 142  
DB 121 WIFINESDDTVSAPARPAVAAA 142  
RESULT 2  
GSPD\_KLEPN STANDARD; PRT; 660 AA.  
ID GSPD\_KLEPN  
AC P15644;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE General secretion pathway protein D precursor (Pullulanase secretion envelope pULD).  
GN PULD.  
OS Klebsiella pneumoniae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Klebsiella.  
OX NCBI\_TaxID=573;  
RN [1]\_TaxID=573;  
RP SEQUENCE FROM N.A. AND SEQUENCE OF 28-42.  
RX MEDLINE=90008916; PubMed=2677007;  
RA D'Entert C., Reys I., Wandersman C., Pugsley A.P.;  
RT "Protein secretion by Gram-negative bacteria. Characterization of two membrane proteins required for pullulanase secretion by Escherichia coli K-12.";  
RL J. Biol. Chem. 264:17462-17468(1989).  
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLLOCATION OF PULLULANASE.  
CC -!- SUBCELLULAR LOCATION: Outer membrane.  
CC -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: M32613; AAA25126.2;  
DR PIR: B34469; B34469.  
DR InterPro: IPR001775; Bac\_GSPD.  
DR InterPro: IPR004846; GSPII/IIIprotein.  
DR InterPro: IPR004845; GSPIIproteinC.  
DR Pfam: PF00263; GSPII\_III; 1.  
DR Pfam: PF03958; GSPII\_III N; 3.  
DR PRINTS: PR00811; BCTERIALGSPD.  
DR PROSITE: PS00875; T2SP\_D; 1.  
KW Transport; Outer membrane; Signal.  
FT SIGNAL 1 27  
FT CHAIN 28 660 GENERAL SECRETION PATHWAY PROTEIN D.  
SQ SEQUENCE 660 AA; 70658 MW; DE25D7C924B85F00 CRC64;

Query Match 1.6%; Score 12; DB 1; Length 660;  
Best Local Similarity 100.0%; Pred. No. 0.0013;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPFLGDIPIVG 736  
DB 559 KVPFLGDIPIVG 570



```

SQ SEQUENCE 445 AA; 49208 MW; 0901DA0D3D42D0E2 CRC64;
Query Match 1.4%; Score 11; DB 1; Length 445;
Best Local Similarity 100.0%; Pred. No. 0.0099;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 725 KVPLGDIPIVI 735
DB 393 KVPLGDIPIVI 403

RESULT 4
PILQ_PSEAE
ID PILQ_PSEAE STANDARD; PRT; 714 AA.
AC P34750;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fimbrial assembly protein pilQ precursor.
GN PILQ OR PA5040.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25102 / PAK;
RX MEDLINE=94049125; PubMed=7901733;
RA Martin P.R., Hobbs M., Free P.D., Jeske Y., Mattick J.S.;
RT "Characterization of pilQ, a new gene required for the biogenesis of
RL type 4 fimbriae in Pseudomonas aeruginosa";
RM Mol. Microbiol. 9:857-868(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RM Nature 406:959-964(2000).
CC -!- FUNCTION: ESSENTIAL FOR THE FORMATION OF PILI. INVOLVED IN THE
CC BIOGENESIS OF TYPE 4 FIMBRIAE PROBABLY BY SERVING AS A "POTHOLE"
CC ALLOWING PASSAGE OF THE FIMBRAE THROUGH THE OUTER MEMBRANE.
CC -!- SUBCELLULAR LOCATION: Outer membrane (Probable).
CC -!- SIMILARITY: BELONGS TO THE PULD/OUTD/EXED/XPSD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L13865; AAA16704.1; ALT_INIT.
CC EMBL; AE004917; AAG08425.1; -.
CC PIR; A83016; A83016.
CC PIR; S37345; S37345.
CC InterPro; IPR001775; Bac_GSPD.
CC InterPro; IPR004846; GSPII/IIIprotein.
CC InterPro; IPR004845; GSPIIproteinC.
CC InterPro; IPR005644; NOLW-like.
CC Pfam; PF00263; GSPII_III; 1.
CC Pfam; PF03958; GSPII_III_N; 1.
CC PRINTS; PR00811; BCTERIALGSPD.
CC PROSITE; PS00875; T2SP_D; 1.
CC PROSITE; PS00875; T2SP_D; 1.
CC TRANSPORT; Protein transport; Outer membrane; Fimbria; Signal;
CC Complete proteome.

```

FT SIGNAL 1 24 POTENTIAL.  
FT CHAIN 25 714 FIMBRIAL ASSEMBLY PROTEIN PILQ.  
FT CONFLICT 391 391 D -> E (IN REF. 1).  
FT CONFLICT 411 411 G -> A (IN REF. 1).  
FT CONFLICT 547 558 LSAMEXTGNCEI -> PVGDGDKQRVV (IN REF. 1).  
SQ SEQUENCE 714 AA; 77378 MW; 9A748EBE826FBAB5 CRC64;  
Query Match 1.4%; Score 11; DB 1; Length 714;  
Best Local Similarity 100.0%; Pred. No. 0.015;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 515 QVMIEARIVEA 525  
| | | | |  
Db 449 QVMIEARIVEA 459  
RESULT 5  
GSPD\_ERWCA STANDARD; PRT; 650 AA.  
ID GSPD\_ERWCA  
AC P31701;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE General secretion pathway protein D precursor (Pectic enzymes  
DE secretion protein outD).  
GN OUTD.  
OS Erwinia carotovora.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Pectobacterium.  
OX NCBI\_TaxID=554;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SGRI193;  
RX MEDLINE=9316842; PubMed=832859;  
RA Reeves P.J., Whitcombe D., Wharam S., Gibson M., Allison G., Bunce N.,  
RA Barrallion R., Douglas P., Mulholland V., Stevens S., Walker S.,  
RA Salmund G.P.C.;  
RT "Molecular cloning and characterization of 13 out genes from Erwinia  
RT carotovora subspecies carotovora; genes encoding members of a general  
RT secretion pathway (GSP) widespread in Gram-negative bacteria."  
RL Mol. Microbiol. 8:443-456(1993).  
RN [2]  
RP IDENTIFICATION OF PROBABLE FRAMESHIFT.  
RA Bairoch A.;  
RL Unpublished observations (FEB 1997).  
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE  
CC EXPORT OF PROTEINS.  
CC MULTIPLE PECTIC ENZYMES.  
CC -!- SUBCELLULAR LOCATION: Outer membrane (Probable).  
CC -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X70049; CAA49645.1; -.  
DR PIR; S32858; S32858.  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSP1/IIIprotein.  
DR InterPro; IPR004845; GSP1/IIIproteinC.  
DR InterPro; IPR005644; NolW-like.  
DR InterPro; IPR003522; SecII\_OMP.  
DR Pfam; PF00263; GSP11\_III; 1.  
DR Pfam; PF03958; GSP11\_III\_N; 3.  
DR PRINTS; PR00811; BCTERIALGSPD.  
DR PROSITE; PS00875; T2SP\_D; 1.  
DR PROSITE; PS00875; T2SP\_D; 1.  
KW Transport; Outer membrane; Signal.  
FT SIGNAL 1 18 POTENTIAL.

FT CHAIN 19 650 GENERAL SECRETION PATHWAY PROTEIN D.  
FT CONFLICT 139 164 ELNDNWRGTGCDYEPANVVMTGRA -> VERQQRVAVDMV  
FT CONFLICT 139 164 RLTCERRRDDWPR (IN REF. 1).  
SQ SEQUENCE 650 AA; 70144 MW; 9A228C369B0E2AFC CRC64;  
Query Match 1.3%; Score 10; DB 1; Length 650;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 725 KVPLLGDIPV 734  
| | | | |  
Db 548 KVPLLGDIPV 557  
RESULT 6  
GSPD\_AERHY STANDARD; PRT; 678 AA.  
ID GSPD\_AERHY  
AC P31780;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE General secretion pathway protein D precursor.  
GN EXED.  
OS Aeromonas hydrophila.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Aeromonadales;  
OC Aeromonadaceae; Aeromonas.  
OX NCBI\_TaxID=644;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Ah65;  
RA Howard S.P.;  
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 408-678 FROM N.A.  
RC STRAIN=Ah65; PubMed=1640836;  
RX MEDLINE=92349963; PubMed=1640836;  
RA Jiang B., Howard S.P.;  
RT "The Aeromonas hydrophila exeE gene, required both for protein  
RT secretion and normal outer membrane biogenesis, is a member of a  
RT general secretion pathway."  
RL Mol. Microbiol. 6:1351-1361(1992).  
RN [2]  
RP FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE  
RP EXPORT OF PROTEINS.  
RP -!- SUBCELLULAR LOCATION: Outer membrane (Probable).  
RP -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.  
RP This SWISS-PROT entry is copyright. It is produced through a collaboration  
RP between the Swiss Institute of Bioinformatics and the EMBL outstation -  
RP the European Bioinformatics Institute. There are no restrictions on its  
RP use by non-profit institutions as long as its content is in no way  
RP modified and this statement is not removed. Usage by and for commercial  
RP entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
RP or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
RP -----  
DR EMBL; X66504; CAA47124.1; -.  
DR PIR; S22668; S22668.  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSP1/IIIprotein.  
DR InterPro; IPR004845; GSP1/IIIproteinC.  
DR InterPro; IPR005644; NolW-like.  
DR Pfam; PF00263; GSP11\_III; 1.  
DR Pfam; PF03958; GSP11\_III\_N; 3.  
DR PRINTS; PR00811; BCTERIALGSPD.  
DR PROSITE; PS00875; T2SP\_D; 1.  
DR PROSITE; PS00875; T2SP\_D; 1.  
KW Transport; Outer membrane; Signal.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 678 GENERAL SECRETION PATHWAY PROTEIN D.  
SQ SEQUENCE 678 AA; 72451 MW; 43B33A28861B0238 CRC64;  
Query Match 1.3%; Score 10; DB 1; Length 678;  
Best Local Similarity 100.0%; Pred. No. 0.15;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPILGDIIPV 734  
DB 565 KVPILGDIIPV 574

RESULT 7  
GSPPD\_AERSA  
ID\_GSPD\_AERSA STANDARD; PRT; 678 AA.  
AC P45778;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE General secretion pathway protein D precursor.  
GN EXED.

OS Aeromonas salmonicida.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Aeromonadales;  
OC Aeromonadaceae; Aeromonas.  
OX NCBI\_TaxID=645;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCIMB 1102;  
RX MEDLINE=95309729; PubMed=7789814;  
RA Karlyshev A.V., Macintyre S.;  
RT "Cloning and study of the genetic organization of the exe gene cluster of Aeromonas salmonicida.";  
RL Gene 158:77-82 (1995).  
CC -!- FUNCTION INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE EXPORT OF PROTEINS.  
CC -!- SUBCELLULAR LOCATION: Outer membrane (Probable).  
CC -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licenses@isb-sib.ch).

DR EMBL; M74011; AAC37020.1; -.  
DR PIR; C40361; C40361.  
DR InterPro; IPR001775; Bac GSPD.  
DR InterPro; IPR004846; GSPII/IIPprotein.  
DR InterPro; IPR004845; GSPIIIproteinC.  
DR InterPro; IPR005644; NclW-like.  
DR InterPro; IPR003522; SecEII\_OMP.  
DR Pfam; PF00263; GSPII\_III; 1.  
DR Pfam; PF03958; GSPII\_III\_N; 2.  
DR PRINTS; PR01337; TYPE3OMGPROT.  
DR PROSITE; PS00875; T2SP\_D; 1.  
DR PLASMIT; TransPort; Protein transport; Outer membrane; Signal.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 678 GENERAL SECRETION PATHWAY PROTEIN D.  
SQ SEQUENCE 678 AA; 67268 MW; CB4921C9BAAB438E CRC64;

Query Match 1.3%; Score 10; DB 1; Length 678;  
Best Local Similarity 100.0%; Pred. No. 0.15;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPILGDIIPV 734  
DB 565 KVPILGDIIPV 574

RESULT 8  
YSKC\_YEREN  
ID\_YSCC\_YEREN STANDARD; PRT; 607 AA.  
AC Q01244;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE YOP proteins translocation protein C precursor.  
GN YOP proteins translocation protein C precursor.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.,  
RA "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474 (1997).  
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE  
CC EXPORT OF PROTEINS (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Outer membrane (Probable).  
CC -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR ENBL; U18997; AAA58122.1; ALT INIT.  
DR ENBL; AE000409; AAC76350.1; ALT\_INIT.  
DR EcoGene; EG12890; gspB.  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSP11/IIIProtein.  
DR InterPro; IPR004845; GSP11proteinc.  
DR InterPro; IPR005644; NolW-like.  
DR InterPro; IPR003522; SecYII-OMPG.  
DR Pfam; PF00263; GSP11\_III; 1.  
DR Pfam; PF03958; GSP11\_III\_N; 3.  
DR PRINTS; PR00811; BCTERIALGSPD.  
DR PRINTS; PR01337; TYPE3OMPGPROF.  
DR PROSITE; PS00875; T2SP D; 1.  
DR TRANSPORT; Outer membrane; Signal; Complete proteome.  
FT SIGNAL 1 23 POTENTIAL.  
FT CHAIN 24 650 PROBABLE GENERAL SECRETION PATHWAY  
FT PROTEIN D.  
FT SEQUENCE 650 AA; 70698 MW; 973459A12A7237B2 CRC64;  
  
Query Match 1.2%; Score 9; DB 1; Length 650;  
Best Local Similarity 100.0%; Pred.No. 1.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 725 KVPLLGDIP 733  
Db 559 KVPLLGDIP 567  
|||||  
  
RESULT 10  
GSPD\_PSEAE STANDARD; PRT; 658 AA.  
AC P35818; Q9H2B2; (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE General secretion pathway protein D precursor.  
GN XCPQ OR PA3105.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=287;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=ATCC 15692 / PAO1;  
RC MEDLINE=95020542; PubMed=7934833;  
RA Akrim M., Bally M., Ball G., Tommassen J., Teerink H., Filloux A.,  
RA Lardunski A.,  
RT "Xcp-mediated protein secretion in Pseudomonas aeruginosa:  
RT identification of two additional genes and evidence for regulation of  
RT xcp gene expression.";  
RL Mol. Microbiol. 10:431-443 (1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PAO1;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
RA Brody R.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.,  
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964 (2000).  
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE  
CC EXPORT OF PROTEINS.  
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).  
CC -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR ENBL; X68594; CAA48582.1; -.  
DR ENBL; AE004734; AAG06493.1; -.  
DR PIR; S39653; S39653.  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSP11/IIIProtein.  
DR InterPro; IPR004845; GSP11proteinc.  
DR InterPro; IPR005644; NolW-like.  
DR Pfam; PF00263; GSP11\_III; 1.  
DR Pfam; PF03958; GSP11\_III\_N; 3.  
DR PRINTS; PR00811; BCTERIALGSPD.  
DR PROSITE; PS00875; T2SP D; 1.  
DR TRANSPORT; Outer membrane; Signal; Complete proteome.  
FT SIGNAL 1 34 POTENTIAL.  
FT CHAIN 35 658 GENERAL SECRETION PATHWAY PROTEIN D.  
FT SEQUENCE 658 AA; 69953 MW; EC2F81FD1A185D50 CRC64;  
  
Query Match 1.2%; Score 9; DB 1; Length 658;  
Best Local Similarity 100.0%; Pred.No. 1.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 725 KVPLLGDIP 733  
Db 569 KVPLLGDIP 577  
|||||  
  
RESULT 11  
GSPD\_VIBCH STANDARD; PRT; 674 AA.  
AC P45779;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE General secretion pathway protein D precursor (Cholera toxin secretion  
DE protein epsD).  
GN EPSD OR VC2733.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=El Tor TRH7000;  
RA Overbye L.J.;  
RT "Organization of the general secretion pathway genes in Vibrio  
RT cholerae.";  
RL Thesis (1994), Michigan State University / East Lansing, U.S.A.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=El Tor N16961 / Serotype O1;  
RX MEDLINE=20406633; PubMed=10953101;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,

RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA Emolaeva M.D., Vamathevan J., Baes S., Qin H., Dragoi I., Sellers P.,  
RA McDonald L., Uterback T., Fieschmann R.D., Nierman W.C., White O.,  
RA Salzberg S.L., Smith H.O., Colwell R.J., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
RT cholerae";  
RL Nature 406:477-483(2000).  
CC -1- FUNCTION: REQUIRED FOR SECRETION OF CHOLERA TOXIN THROUGH THE  
CC -1- OUTER MEMBRANE.  
CC  
CC -1- SUBCELLULAR LOCATION: Outer membrane (Potential).  
CC  
CC -1- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.  
CC  
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
CC FRAMESHIFT (POSSIBLY NATURAL) IN POSITION 55.  
CC  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
CC EMBL; L33796; AAA59785.1; -;  
CC EMBL; A8004338; -; NOT\_ANNOTATED\_CDS.  
CC TIGR; VC2733; -;  
CC InterPro; IPR001775; Bac\_GSPD.  
CC InterPro; IPR004846; GSPiI/IIprotein.  
CC InterPro; IPR004845; GSPiIproteinC.  
CC InterPro; IPR005644; NOLW-like.  
CC InterPro; IPR00263; GSPiI\_III.1.  
CC Pfam; PF03958; GSPiI\_III\_N; 3.  
CC Pfam; PF03958; GSPiI\_III\_N; 3.  
CC PRINTS; PR00811; BCTERIALGSPD.  
CC PROSITE; PS00875; T2SP.D; 1.  
CC Transprot; Outer membrane; Signal; Complete proteome.  
CC  
CC SIGNAL 1 24  
CC POTENTIAL.  
CC  
CC FT CHAIN 25 674 GENERAL SECRETION PATHWAY PROTEIN D.  
FT CONFLICT 89 89 V -> A (IN REF. 1).  
FT CONFLICT 144 144 R -> P (IN REF. 1).  
FT CONFLICT 674 674 3D77B891A5956223 CRC64;  
SQ  
Query Match 1.2%; Score 9; DB 1; Length 674;  
Best Local Similarity 100.0%; Pred.No. 1.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 725 KVPLLGDIP 733  
Db 575 KVPLLGDIP 593  
RESULT 12  
ID\_HRPH\_PSEY STANDARD; PRT; 701 AA.  
AC Q01723; 1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Hypersensitivity response secretion protein hrph precursor.  
GN HRPH.  
OS Pseudomonas syringae (pv. syringae).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
CX NCBI\_TaxID=321;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=61;  
RX MEDLINE=93015750; PubMed=1400238;  
RA Huang H.-C., He S.Y., Bauer D.W., Collmer A.;  
RT "The Pseudomonas syringae pv. syringae 61 hrph product, an envelope  
RT protein required for elicitation of the hypersensitive response in  
RT plants.";  
RL J. Bacteriol. 174:6878-6885(1992).  
RN [2]

RP REVISIONS.  
RA Deng W.-L.;  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1-52 FROM N.A.  
RC STRAIN=61;  
RX MEDLINE=96025090; PubMed=7579617;  
RA Huang H.-C., Lin R.H., Chang C.J., Collmer A., Deng W.-L.;  
RT "The complete hrp gene cluster of Pseudomonas syringae pv. syringae  
RT 61 includes two blocks of genes required for harpinPss secretion that  
RT are arranged colinearly with Yersinia ysc homologs.";  
RL Mol. Plant Microbe Interact. 8:733-746(1995).  
CC -1- FUNCTION: INVOLVED IN THE SECRETION OF A PROTEINACEOUS ELICITOR OF  
CC THE HYPERSENSITIVITY RESPONSE IN PLANTS.  
CC  
CC -1- SUBCELLULAR LOCATION: Outer membrane.  
CC  
CC -1- SIMILARITY: BELONGS TO THE PULD/OUTD/EXED/XPSD FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
CC EMBL; L01064; AAC05014.1; -;  
CC EMBL; U25813; AAB05085.1; -;  
CC InterPro; IPR001775; Bac\_GSPD.  
CC InterPro; IPR004846; GSPiI/IIprotein.  
CC InterPro; IPR004845; GSPiIproteinC.  
CC InterPro; IPR005644; NOLW-like.  
CC InterPro; IPR003522; GSPiI\_III\_N; 2.  
CC Pfam; PF00263; GSPiI\_III\_N; 2.  
CC Pfam; PF03958; GSPiI\_III\_N; 2.  
CC PRINTS; PR00811; BCTERIALGSPD.  
CC PRINTS; PR01337; TYPE3OMGPROT.  
CC PROSITE; PS00875; T2SP.D; FALSE NEG.  
CC Protein transport; Transport; Signal; Outer membrane; Translocation;  
CC Hypersensitive response.  
CC  
CC SIGNAL 1 21  
CC POTENTIAL.  
CC  
CC FT CHAIN 22 701  
CC HYPERSENSITIVITY RESPONSE SECRETION  
CC PROTEIN HRPH.  
SQ  
Query Match 1.2%; Score 9; DB 1; Length 701;  
Best Local Similarity 100.0%; Pred.No. 1.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 725 KVPLLGDIP 733  
Db 502 KVPLLGDIP 510  
RESULT 13  
ID\_GSQD\_ERWCH STANDARD; PRT; 710 AA.  
AC Q01565;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE General secretion pathway protein D precursor (Pectic enzymes  
DE secretion protein outd).  
GN OUTD.  
OS Erwinia chrysanthemi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Pectobacterium.  
CX NCBI\_TaxID=556;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=3937;  
RX MEDLINE=93086427; PubMed=1453958;  
RA Condemine G., Dorel C., Hugouvieux-Cotte-Pattat N., Robert-Baudouy J.;  
RT "Some of the out genes involved in the secretion of pectate lyases in

```

RT Erwinia chrysanthemi are regulated by kdgR.
RL Mol. Microbiol. 6:3199-3211(1992).
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLLOCATION OF THE
CC MULTIPLE PECTIC ENZYMES.
CC -!- SUBCELLULAR LOCATION: Outer membrane (Probable).
CC -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X65265; CAA46370.1; -
CC PIR; S28014;
CC InterPro; IPR001775; Bac GSPD.
CC InterPro; IPR004846; GSPFII/IIPprotein.
CC InterPro; IPR004845; GSPFIIproteinc.
CC Pfam; PF00263; GSPFII_III; 1.
CC Pfam; PF03958; GSPFII_III_N; 3.
CC PRINTS; PR00811; BCTERIALGSPD.
CC PROSITE; PS00875; T2SP_D; 1.
CC TRANSPORT; Outer membrane; Signal.
CC SIGNAL 1 27 POTENTIAL.
CC FT CHAIN 28 710 GENERAL SECRETION PATHWAY PROTEIN D.
CC FT DOMAIN 288 353 GLY/SER-RICH.
CC SQ SEQUENCE 710 AA; 76213 MW; 156E84CC50CD54FA CRC64;

Query Match 1.2%; Score 9; DB 1; Length 710;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPILGDIP 733
DB 612 KVPILGDIP 620

RESULT 14
GSPD ERWCH
ID GSPD ERWCH STANDARD; PRT; 712 AA.
AC P31700;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE General secretion pathway protein D precursor (Pectic enzymes
DE secretion protein outD).
GN OUTD.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EC16;
RX MEDLINE=93054355; PubMed=1429461;
RA Lindeberg M., Collmer A.;
RT "Analysis of eight out genes in a cluster required for pectic enzyme
RT secretion by Erwinia chrysanthemi: sequence comparison with secretion
RT genes from other Gram-negative bacteria.";
RL J. Bacteriol. 174:7385-7397(1992).
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLLOCATION OF THE
CC MULTIPLE PECTIC ENZYMES.
CC -!- SUBCELLULAR LOCATION: Outer membrane (Probable).
CC -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L02214; AAA24831.1; -
CC InterPro; IPR001775; Bac GSPD.
CC InterPro; IPR004846; GSPFII/IIPprotein.
CC InterPro; IPR004845; GSPFIIproteinc.
CC Pfam; PF00263; GSPFII_III; 1.
CC Pfam; PF03958; GSPFII_III_N; 3.
CC PRINTS; PR00811; BCTERIALGSPD.
CC PROSITE; PS00875; T2SP_D; 1.
CC TRANSPORT; Outer membrane; Signal.
CC SIGNAL 1 27 POTENTIAL.
CC FT CHAIN 28 712 GENERAL SECRETION PATHWAY PROTEIN D.
CC FT DOMAIN 288 353 GLY/SER-RICH.
CC SQ SEQUENCE 712 AA; 76478 MW; 8A065D9ADA24889 CRC64;

Query Match 1.2%; Score 9; DB 1; Length 712;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPILGDIP 733
DB 614 KVPILGDIP 622

RESULT 15
LHA3_RHOPA
ID LHA3_RHOPA STANDARD; PRT; 65 AA.
AC P35103;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Light-harvesting protein B-800-850, alpha chain C (Antenna pigment
DE protein, alpha chain C) (LH II-C alpha).
GN PUCAC.
OS Rhodospseudomonas palustris.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Rhodospseudomonas.
OX NCBI_TaxID=1076;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-9.
RC STRAIN=IE5;
RX MEDLINE=89356600; PubMed=2670551;
RA Tadtos M.H., Waterkamp K.;
RT "Multiple copies of the coding regions for the light-harvesting
RT B800-850 alpha- and beta-polypeptides are present in the
RT Rhodospseudomonas palustris genome.";
RL EMBO J. 8:1303-1308(1989).
CC -!- FUNCTION: ANTENNA COMPLEXES ARE LIGHT-HARVESTING SYSTEMS, WHICH
CC -!- SUBUNIT: THE CORE COMPLEX IS FORMED BY DIFFERENT ALPHA AND BETA
CC CHAINS, BINDING BACTERIOCHLOROPHYLL MOLECULES, AND ARRANGED
CC MOST PROBABLY IN TETRAMERIC STRUCTURES DISPOSED AROUND THE
CC REACTION CENTER. THE NON-PIGMENTED GAMMA CHAINS MAY CONSTITUTE
CC ADDITIONAL COMPONENTS.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X64958; CAA46122.1; -
CC HSSP; P26789; 1KZU.
CC InterPro; IPR000066; Antenna a/b.
CC InterPro; IPR002361; Antenna a.

```

DR Pfam; PF00556; LHC; 1.  
 DR PROSITE; PS00968; ANTENNA\_COMP\_ALPHA; 1.  
 KW Antenna complex; light-harvesting polypeptide; Transmembrane;  
 KW Magnesium; Bacteriochlorophyll; Inner membrane.  
 FT DOMAIN 1 11 CYTOPLASMIC (POTENTIAL).  
 FT TRANSFAM 12 35 POTENTIAL.  
 FT DOMAIN 36 65 PERIPLASMIC (POTENTIAL).  
 FT METAL 31 31 MAGNESIUM (BACTERIOCHLOROPHYLL AXIAL LIGAND) (POTENTIAL).  
 FT SEQUENCE 65 AA; 6557 MW; C2E25F9C2A2D0BFD CRC64;  
 Query Match 1.0%; Score 8; DB 1; Length 65;  
 Best Local Similarity 100.0%; Pred.No. 2.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 141 AAPAAPAK 148  
 DB 57 AAPAAPAK 64  
 ID YC91\_MYCTU STANDARD; PRT; 111 AA.  
 AC Q10617;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein Rv1291c.  
 GN Rv1291C OR MT1330 OR MTCY373.10C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weyman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO M.TUBERCULOSIS Rv1271C.  
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
 CC FRAMESHIFT IN POSITION 56.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z73419; CAA97761.1; -;  
 CC EMBL; AE007007; AK45591.1; ALT\_FRAME.  
 DR EMBL; Z73419; CAA97761.1; -;  
 DR EMBL; AE007007; AK45591.1; ALT\_FRAME.

DR PIR; G70772; G70772.  
 DR TIGR; MT1330; -;  
 DR Tuberculin; Rv1291c; -;  
 DR Pfam; PF05305; DUF732; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 111 AA; 11025 MW; 45585EA14834BB41 CRC64;  
 Query Match 1.0%; Score 8; DB 1; Length 111;  
 Best Local Similarity 100.0%; Pred.No. 3.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 242 LAALGFAG 249  
 DB 22 LAALGFAG 29  
 ID HOFQ\_ECOLI STANDARD; PRT; 412 AA.  
 AC P34749;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein transport protein hofQ precursor.  
 GN HOFQ OR HOFQ OR B3391.  
 DE HOFQ OR HOFQ OR B3391.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12."  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP PRELIMINARY SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=92105021; PubMed=1309529;  
 RA Lobner-Olesen A., Boye E., Marinus M.G.;  
 RT "Identification of the gene (arok) encoding shikimic acid kinase I of  
 RT Escherichia coli."  
 RL J. Bacteriol. 174:525-529(1992).  
 RN [3]  
 RP IDENTIFICATION  
 RX MEDLINE=94049125; PubMed=7901733;  
 RA Martin P.R., Hobbs M., Free P.D., Jeske Y., Mattick J.S.;  
 RT "Characterization of pilQ, a new gene required for the biogenesis of  
 RT type 4 fimbriae in Pseudomonas aeruginosa."  
 RL Mol. Microbiol. 9:857-868(1993).  
 CC -1- SUBCELLULAR LOCATION: Outer membrane (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE PULD/OUTP/EXED/XPSD FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U18997; AAA58188.1; -;  
 CC EMBL; AE000414; AAC76416.1; -;  
 CC EMBL; Z19601; -; NOT ANNOTATED\_CDS.  
 DR EMBL; Z19601; -; NOT ANNOTATED\_CDS.  
 DR PIR; B65134; B65134; -;  
 DR EcoGene; EG12113; hofQ.  
 DR InterPro; IPR001775; Bac\_GSPD.  
 DR InterPro; IPR004846; GSPFII/IIprotein.  
 DR InterPro; IPR004845; GSPFIIproteinC.

DR InterPro; IPR005644; NOLW-like.  
DR InterPro; IPR003522; SecIII\_OMP.  
DR Pfam; PF00263; GSP11\_III\_1  
DR Pfam; PF03958; GSP11\_III\_N; 1.  
DR PRINTS; PRO0811; BCTERIALGSPD.  
DR PRINTS; PRO1337; TYPE3OMGPROT.  
DR PROSITE; PS00875; T2SP\_D; 1.  
KW Transport; Outer membrane; Signal; Complete proteome.  
FT SIGNAL 1 18  
FT CHAIN 19 412  
SQ SEQUENCE 412 AA; 44716 MW; 99FEAE2606682E5 CRC64;  
Query Match 1.0%; Score 8; DB 1; Length 412;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 726 VPLGSDIP 733  
DB 375 VPLGSDIP 382  
RESULT 18  
ID HEM2 SELMA STANDARD; PRT; 417 AA.  
AC P45623;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Delta-aminolevulinic acid dehydratase, chloroplast precursor  
DE (EC 4.2.1.24) (Porphobilinogen synthase) (ALADH) (ALAD).  
GN ALA2.  
OS Selaginella martensii (Martens's spike moss).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Lycopodiophyta; Isoetes; Selaginellales; Selaginellaceae;  
OC Selaginella.  
OX NCBI\_TaxID=3247;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Solbach M., Schneider-Poetsch H.A.W.;  
RL Submitted (Sep-1993) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: 2 5-aminolevulinate = porphobilinogen + 2 H(2)O.  
CC -!- COFACTOR: MAGNESIUM.  
CC -!- PATHWAY: Porphyrin biosynthesis; second step.  
CC -!- SUBUNIT: Homotrimer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Chloroplast.  
CC -!- SIMILARITY: BELONGS TO THE ALADH FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X75043; CAA52955.1; -  
CC HSP; F15002; I84E.  
DR InterPro; IPR001731; AlaD\_dehydratase.  
DR Pfam; PF00490; ALAD; 1.  
DR PRINTS; PRO1144; DALDHDTASE.  
DR ProDom; PD002304; AlaD\_dehydratase; 1.  
DR PROSITE; PS00169; D\_ALA\_DEHYDRATASE; 1.  
KW Porphyrin biosynthesis; Chlorophyll biosynthesis; Lyase; Magnesium;  
KW Chloroplast; transit peptide.  
FT TRANSIT 1 40  
FT CHAIN 41 417  
FT DOMAIN 207 225  
FT ACT SITE 339 339  
FT BY SIMILARITY.  
SQ SEQUENCE 417 AA; 45182 MW; 036E57A607886759 CRC64;  
Query Match 1.0%; Score 8; DB 1; Length 417;  
Best Local Similarity 100.0%; Pred. No. 10;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 141 AAPAPAK 148  
DB 70 AAPAPAK 77  
RESULT 19  
ID NUCLEOLIN STANDARD; PRT; 706 AA.  
AC P09405; Q61991;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Nucleolin (Protein C23).  
GN NCL OR NUC.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C;  
RX MEDLINE=88316930; PubMed=3137346;  
RA Bourbon H., Lapeyre B., Amalric F.;  
RT "Structure of the mouse nucleolin gene. The complete sequence reveals  
RT that each RNA binding domain is encoded by two independent exons.";  
RL J. Mol. Biol. 200:627-638(1988).  
RN [2]  
RP SEQUENCE OF 1-44 FROM N.A.  
RX MEDLINE=89121496; PubMed=2906027;  
RA Bourbon H.M., Prudhomme M., Amalric F.;  
RT "Sequence and structure of the nucleolin promoter in rodents:  
RT characterization of a strikingly conserved CpG island.";  
RL Gene 68:73-84(1988).  
RN [3]  
RP SEQUENCE OF 1-23.  
RX MEDLINE=91317840; PubMed=1860869;  
RA Pasternack M.S., Bleier K.J., McInerney T.N.;  
RT "Granzyme A binding to target cell proteins. Granzyme A binds to and  
RT cleaves nucleolin in vitro.";  
RL J. Biol. Chem. 266:14703-14708(1991).  
CC -!- FUNCTION: NUCLEOLIN IS THE MAJOR NUCLEOLAR PROTEIN OF GROWING  
CC EUKARYOTIC CELLS. IT IS FOUND ASSOCIATED WITH INTRANUCLEOLAR  
CC CHROMATIN AND PRERIBOSOMAL PARTICLES. IT INDUCES CHROMATIN  
CC DECONDENSATION BY BINDING TO HISTONE H1. IT IS THOUGHT TO PLAY A  
CC ROLE IN PRE-RRNA TRANSCRIPTION AND RIBOSOME ASSEMBLY.  
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.  
CC -!- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X07699; CAA30538.1; -  
CC EMBL; M22089; AAA39841.1; -  
DR PIR; A28958; DNMS.  
DR HSP; P09651; 1HAI.  
DR MGD; MGI:97286; Ncl.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF00076; rrm; 4.  
DR SMART; SM00360; RRM; 4.  
DR PROSITE; PS00102; RRM; 4.  
DR PROSITE; PS00030; RRM\_RNP\_1; 3.  
KW Nuclear protein; Phosphorylation; Methylation; DNA-binding; Repeat;  
KW RNA-binding.  
FT INIT MET 0 0  
FT DOMAIN 142 169  
FT DOMAIN 189 214  
FT ASP/GLU-RICH (ACIDIC).  
FT ASP/GLU-RICH (ACIDIC).



```

FT DOMAIN 240 272 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 308 384 RNA-BINDING (RRM) 1.
FT DOMAIN 394 467 RNA-BINDING (RRM) 2.
FT DOMAIN 486 560 RNA-BINDING (RRM) 3.
FT DOMAIN 568 643 RNA-BINDING (RRM) 4.
FT DOMAIN 645 696 ARG/GLY/PHE-RICH.
FT MOD_RES 144 144 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 156 156 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 188 188 PHOSPHORYLATION (BY SIMILARITY).
FT DOMAIN 57 134 8 X 8 AA TANDEM REPEATS OF X-T-P-X-K-K-X-X.
FT REPEAT 57 64 1.
FT REPEAT 74 81 2.
FT REPEAT 82 89 3.
FT REPEAT 90 97 4.
FT REPEAT 98 103 5 (INCOMPLETE).
FT REPEAT 104 111 6.
FT REPEAT 119 126 7.
FT REPEAT 127 134 8.
SQ SEQUENCE 706 AA; 76592 MW; 89505EE39C89F832 CRC64;

Query Match 1.0%; Score 8; DB 1; Length 706;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 140 KAAPAAPA 147
Db 180 KAAPAAPA 187
|||||

RESULT 20
NUCL_RAT
ID NUCL_RAT STANDARD; PRT; 712 AA.
AC P13383;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nucleolin (Protein C23).
GN NCL OR NUC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90269607; PubMed=2347493;
RA Bourbon H.-M., Amalric F.;
RT "Nucleolin gene organization in rodents: highly conserved sequences within three of the 13 introns.";
RL Gene 88:187-196(1990).
RN [2]
RP SEQUENCE OF 1-44 FROM N.A.
RX MEDLINE=89121495; PubMed=2906027;
RA Bourbon H.-M., Prudhomme M., Amalric F.;
RT "Sequence and structure of the nucleolin promoter in rodents: characterization of a strikingly conserved CpG island.";
RL Gene 68:73-84(1988).
CC -1- FUNCTION: NUCLEOLIN IS THE MAJOR NUCLEOLAR PROTEIN OF GROWING EUKARYOTIC CELLS. IT IS FOUND ASSOCIATED WITH INTRANUCLEOLAR CHROMATIN AND PRERIBOSOMAL PARTICLES. IT INDUCES CHROMATIN DECONDENSATION BY BINDING TO HISTONE H1. IT IS THOUGHT TO PLAY A ROLE IN PRE-RRNA TRANSCRIPTION AND RIBOSOME ASSEMBLY.
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -1- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license@sb-sib.ch. (See http://www.isb-sib.ch/announce/ or send an email to license@sb-sib.ch).

```

```

DR EMBL; M55022; AAA41732.1; -.
DR EMBL; M55015; AAA41732.1; JOINED.
DR EMBL; M55017; AAA41732.1; JOINED.
DR EMBL; M55020; AAA41732.1; JOINED.
DR EMBL; M22090; AAA41732.1; -.
DR HSP; P09651; IHA1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 4.
DR SMART; SM00360; RRM; 4.
DR PROSITE; PS0102; RRM; 4.
DR PROSITE; PS0030; RRM_RNP_1; 3.
KW Nuclear protein; Phosphorylation; Methylation; DNA-binding; Repeat; RNA-binding.
KW INIT MET 0
FT DOMAIN 142 167 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 187 215 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 241 274 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 310 386 RNA-BINDING (RRM) 1.
FT DOMAIN 396 469 RNA-BINDING (RRM) 2.
FT DOMAIN 488 562 RNA-BINDING (RRM) 3.
FT DOMAIN 574 649 RNA-BINDING (RRM) 4.
FT DOMAIN 651 702 ARG/GLY/PHE-RICH.
FT DOMAIN 57 134 8 X 8 AA TANDEM REPEATS OF X-T-P-X-K-K-X-X.
FT REPEAT 57 64 1.
FT REPEAT 74 81 2.
FT REPEAT 82 89 3.
FT REPEAT 90 97 4.
FT REPEAT 98 103 5 (INCOMPLETE).
FT REPEAT 104 111 6.
FT REPEAT 119 126 7.
FT REPEAT 127 134 8.
SQ SEQUENCE 712 AA; 77016 MW; 68774A214E550F90 CRC64;

Query Match 1.0%; Score 8; DB 1; Length 712;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 140 KAAPAAPA 147
Db 178 KAAPAAPA 185
|||||

RESULT 21
DMSA_RHOCA
ID DMSA_RHOCA STANDARD; PRT; 823 AA.
AC Q52675; P72249;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Dimethyl sulfoxide/trimethylamine N-oxide reductase precursor (EC 1.7.2.3) (DMSO reductase) (DMSOR).
GN DORA.
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Rhodobacter.
OX NCBI_TaxId=1061;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 43-59.
RC STRAIN=DSM 938 / 37b4;
RX MEDLINE=97008997; PubMed=8856102;
RA Shaw A.L., Hanson G.R., McEwan A.G.;
RT "Cloning and sequence analysis of the dimethylsulfoxide reductase structural gene from Rhodobacter capsulatus.";
RL Biochim. Biophys. Acta 1276:176-180(1996).
RN [2]
RP REVISIONS.
RA Shaw A.L., McEwan A.G.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 938 / 37b4;
RX MEDLINE=97045989; PubMed=8890911;

```

RA Knaeblein J., Mann K., Ehler S., Fonstein M., Huber R., Schneider F.;  
RT "Isolation, cloning, sequence analysis and localization of the operon  
RT encoding dimethyl sulfoxide/trimethylamine N-oxide reductase from  
RT Rhodobacter capsulatus.";  
RL J. Mol. Biol. 263:40-52(1996).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (1.88 ANGSTROMS).  
RC STRAIN=DSM 938 / 3764;  
RX MEDLINE=97045990; PubMed=8990912;  
RA Schneider F., Lowe J., Huber R., Schindelin H., Kisker C.,  
RA Knaeblein J.;  
RT "Crystal structure of dimethyl sulfoxide reductase from Rhodobacter  
RT capsulatus at 1.88-A resolution.";  
RL J. Mol. Biol. 263:53-69(1996).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
RC STRAIN=H123;  
RX MEDLINE=20296793; PubMed=10835270;  
RA McAlpine A.S., Shaw A.L., Bailey S.;  
RT "Molybdenum active centre of DMSO reductase from Rhodobacter  
RT capsulatus: crystal structure of the oxidised enzyme at 1.82-A  
RT resolution and the dithionite-reduced enzyme at 2.8-A resolution.";  
RL J. Biol. Inorg. Chem. 2:690-700(1997).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
RC STRAIN=H123;  
RX MEDLINE=20442077; PubMed=10985771;  
RA Bray R.C., Adams B., Smith A.T., Bennett B., Bailey S.;  
RT "Reversible dissociation of thiolate ligands from molybdenum in an  
RT enzyme of the dimethyl sulfoxide reductase family.";  
RL Biochemistry 38:11258-11269(2000).  
CC -!- FUNCTION: Terminal reductase during anaerobic growth on  
CC various sulfoxide and n-oxide compounds.  
CC -!- CATALYTIC ACTIVITY: Reduces various N-oxide and sulfoxide  
CC compounds including trimethylamine N-oxide.  
CC -!- COFACTOR: Binds 1 molybdenum ion and 2 molybdopterin guanine  
CC dinucleotide (MGB) groups per subunit.  
CC -!- SUBUNIT: Monomer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Periplasmic.  
CC -!- PTM: Predicted to be exported by the Tat system. The position of  
CC the signal peptide cleavage has been experimentally proven.  
CC -!- SIMILARITY: BELONGS TO THE PROKARYOTIC MOLYBDOPTERIN-CONTAINING  
CC OXIDOREDUCTASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U49506; AAD13674.1; -;  
CC EMBL; X95407; CAA54689.1; ALT\_INIT.  
CC PDB; 1DMR; 18-MAR-98.  
CC PDB; 2DMR; 18-MAR-98.  
CC PDB; 3DMR; 18-MAR-98.  
CC PDB; 4DMR; 18-MAR-98.  
CC PDB; 1DMS; 01-JUL-98.  
CC PDB; 1E18; 21-MAY-01.  
CC PDB; 1ESV; 03-NOV-00.  
CC PDB; 1E60; 25-AUG-00.  
CC PDB; 1E61; 03-NOV-00.  
CC InterPro; IPR006658; Bisc.  
CC InterPro; IPR006657; Mol\_dinuc\_bind.

DR InterPro; IPR006656; Molybdopterin.  
DR InterPro; IPR006655; Prok\_Mboxied.  
DR Pfam; PF00384; molybdopterin; 1.  
DR Pfam; PF01568; Molydop\_binding; 1.  
DR TIGRFAMs; TIGR00509; bisc\_fam; 1.  
DR PROSITE; PS00551; MOLYBDOPTERIN\_PROK\_1; FALSE\_NEG.  
DR PROSITE; PS00490; MOLYBDOPTERIN\_PROK\_2; 1.  
DR PROSITE; PS00932; MOLYBDOPTERIN\_PROK\_3; FALSE\_NEG.  
KW Oxidoreductase; Molybdenum; Metal-binding; Peptidase; Signal;  
KW 3D-structure.  
FT SIGNAL 1 42 TAT-TYPE SIGNAL.  
FT CHAIN 43 823 DIMETHYL SULFOXIDE/TRIMETHYLAMINE N-OXIDE  
FT REDUCTASE.  
FT  
FT ACT SITE 189 189  
FT METAL 189 189  
FT CONFLICT 33 33  
FT CONFLICT 81 81  
FT CONFLICT 85 85  
FT CONFLICT 149 149  
FT CONFLICT 276 278  
FT CONFLICT 322 322  
FT CONFLICT 336 337  
FT CONFLICT 354 354  
FT CONFLICT 354 354  
FT CONFLICT 410 410  
FT CONFLICT 416 416  
FT CONFLICT 498 498  
FT CONFLICT 568 568  
FT CONFLICT 597 597  
FT CONFLICT 769 769  
FT STRAND 48 54  
FT TURN 55 56  
FT STRAND 57 64  
FT TURN 65 66  
FT STRAND 67 73  
FT TURN 75 76  
FT TURN 82 83  
FT HELIX 84 92  
FT TURN 94 95  
FT STRAND 96 96  
FT STRAND 101 103  
FT HELIX 104 109  
FT HELIX 110 112  
FT HELIX 115 117  
FT TURN 118 119  
FT STRAND 123 125  
FT HELIX 128 146  
FT HELIX 148 150  
FT STRAND 151 152  
FT HELIX 167 178  
FT TURN 179 179  
FT STRAND 182 185  
FT TURN 191 192  
FT HELIX 193 200  
FT TURN 201 201  
FT TURN 206 207  
FT HELIX 213 219  
FT STRAND 222 226  
FT HELIX 230 233  
FT TURN 234 235  
FT STRAND 244 244  
FT TURN 245 255  
FT TURN 256 256  
FT STRAND 258 262  
FT STRAND 265 265  
FT HELIX 268 273  
FT TURN 274 274  
FT STRAND 276 278  
FT TURN 282 283  
FT HELIX 285 298  
FT TURN 299 300  
FT HELIX 304 310

```

FT STRAND 311 312
FT HELIX 314 321
FT TURN 322 326
FT HELIX 332 339
FT HELIX 343 355
FT STRAND 358 362
FT HELIX 365 367
FT TURN 370 372
FT HELIX 373 385
FT TURN 387 387
FT TURN 389 390
FT TURN 392 393
FT STRAND 395 397
FT TURN 400 401
FT TURN 403 406
FT HELIX 422 424
FT STRAND 438 440
FT HELIX 441 443
FT TURN 444 449
FT TURN 451 452
FT STRAND 454 457
FT TURN 458 459
FT STRAND 460 463
FT STRAND 469 473
FT HELIX 477 480
FT HELIX 484 490

Query Match 1.0%; Score 8; DB 1; Length 823;
Best Local Similarity 100.0%; Pred.No.18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 391 WDAQALDIV 398
Db 128 WDAQALDIV 135

RESULT 22
KAB7 SCHPO STANDARD; PRT; 836 AA.
AC Q09815;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable serine/threonine-protein kinase C16C9.07 (BC 2.7.1.1.-).
GN SPAC16C9.07 OR SPAC2G11.01.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed11859350;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gattolisi S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Welter J., Vanstreels E., Rieger M., Schaefer M., Muller-Auer S.,
RA Gabel C., Fuchs M., Fritsch C., Holzer E., Mostl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Fumelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

```

```

RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakowski G.V., Usery D., Barrell B.G., Nurse P.;
RL "The genome sequence of Schizosaccharomyces pombe.";
RT Nature 415:871-880(2002).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MNB/DYRK SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z54366; CAA91195.1; -.
CC EMBL; Z54354; CAA91166.1; -.
CC HSSP; P24941; 1AQ1.
CC GenDB SPombe; SPAC16C9.07; -.
CC InterPro; IPR000719; Prot_Kinase.
CC InterPro; IPR002290; Ser_Thr_kinase.
CC Pfam; PF00069; Pkinase_I.
CC ProDom; PD000001; Prot_kinase; 1.
CC SMART; SM00220; S_TKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
CC KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;
KW ATP-binding.
FT DOMAIN 518 814 PROTEIN KINASE.
FT NP_BIND 524 532 ATP (BY SIMILARITY).
FT BINDING 547 547 ATP (BY SIMILARITY).
FT ACT_SITE 644 644 BY SIMILARITY.
SQ SEQUENCE 836 AA; 94231 MW; 2DCD1E965FC66A80 CRC64;

Query Match 1.0%; Score 8; DB 1; Length 836;
Best Local Similarity 100.0%; Pred.No.18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 604 LSKTKTLA 611
Db 180 LSKTKTLA 187

RESULT 23
YEEJ ECOLI STANDARD; PRT; 2358 AA.
AC P76347; P94750;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yeej.
GN YEEJ OR B1978.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;

```

```

RX MEDLINE=97251350; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,
RA Saito N., Sempel G., Seki Y., Sivasundaram S., Tagami H.,
RA Takeda J., Takenoto K., Wada C., Yanamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
PL DNA Res. 3:379-392(1996).
CC -!- SIMILARITY: CONTAINS 13 Big-1 domains.
CC -!- SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000289; AAC75042.1; ALT_INIT.
CC EMBL; D90837; BAA15800.1; -.
CC EMBL; D90836; BAA15799.1; ALT_INIT.
CC EcoGene; EG13378; yeeJ.
CC InterPro; IPR003344; Big_1.
CC InterPro; IPR003535; Intimin.
CC InterPro; IPR002482; LysM.
CC InterPro; IPR000601; PKD_domain.
CC Pfam; PF02369; Big_1; 13.
CC PRINTS; PR01369; INTIMIN.
CC SMART; SM00634; BID_1; 13.
CC SMART; SM00257; LysM; 1.
CC SMART; SM00089; PKD; 6.
CC KW Hypothetical protein; Repeat; Complete proteome.
FT DOMAIN 738 834
FT DOMAIN 840 929
FT DOMAIN 931 1033
FT DOMAIN 1042 1132
FT DOMAIN 1134 1236
FT DOMAIN 1245 1335
FT DOMAIN 1337 1439
FT DOMAIN 1448 1539
FT DOMAIN 1548 1652
FT DOMAIN 1653 1750
FT DOMAIN 1751 1855
FT DOMAIN 1856 1957
FT DOMAIN 1963 2056
FT DOMAIN 2065 2156
FT DOMAIN 2157 2252
FT DOMAIN 2254 2355
FT DOMAIN 2358 AA; 248599 MW; 232249750Bf631ED CRC64;
SQ SEQUENCE 2358 AA; 248599 MW; 232249750Bf631ED CRC64;

Query Match 1.0%; Score 8; DB 1; Length 2358;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 DELDVPAQ 514
DB 113 DELDVPAQ 120

RESULT 24
VEEJ_ECO57
ID VEEJ_ECO57 STANDARD; PRT; 2660 AA.
AC Q8X8V7; Q8X2B9; Q8X2C0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yeeJ.
GN Z3135 OR ECS2775/ECS2776.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;

```

```

RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / EDL933 / ATCC 700927;
RC MEDLINE=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., Burland V., Mau B., Glaesner J.D.,
RA Rose D.J., Maynew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takani H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -!- SIMILARITY: CONTAINS 16 Big-1 domains.
CC -!- SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1315.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE005423; AAG57041.1; -.
CC EMBL; AP002559; BAB36198.1; ALT_FRAME.
CC EMBL; AP002559; BAB36199.1; ALT_FRAME.
CC InterPro; IPR003344; Big_1.
CC InterPro; IPR003535; Intimin.
CC InterPro; IPR000601; PKD_domain.
CC Pfam; PF02369; Big_1; 16.
CC PRINTS; PR01369; INTIMIN.
CC SMART; SM00634; BID_1; 16.
CC SMART; SM00089; PKD; 8.
CC KW Hypothetical protein; Repeat; Complete proteome.
FT DOMAIN 738 834
FT DOMAIN 840 929
FT DOMAIN 931 1033
FT DOMAIN 1042 1132
FT DOMAIN 1134 1236
FT DOMAIN 1245 1335
FT DOMAIN 1337 1439
FT DOMAIN 1448 1539
FT DOMAIN 1548 1652
FT DOMAIN 1653 1750
FT DOMAIN 1751 1855
FT DOMAIN 1856 1957
FT DOMAIN 1963 2056
FT DOMAIN 2065 2156
FT DOMAIN 2157 2252
FT DOMAIN 2254 2355
FT DOMAIN 2358 AA; 280062 MW; Q1EB92A08F5C09D2 CRC64;
SQ SEQUENCE 2660 AA; 280062 MW; Q1EB92A08F5C09D2 CRC64;

Query Match 1.0%; Score 8; DB 1; Length 2660;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 DELDVPAQ 514
DB 113 DELDVPAQ 120

```

```

RESULT 25
DYL4 CHLRE
ID DYL4_CHLRE STANDARD; PRT; 129 AA.
AC Q39591;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Dynein 14 kDa light chain, flagellar outer arm.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474 (1997).
RP SEQUENCE FROM N.A., AND SEQUENCE OF 15-32 AND 58-65.
RC STRAIN=1132D;
RX MEDLINE=96198089; PubMed=8626422;
RA Patel-King R.S., Benashski S.E., Harrison A., King S.M.;
RT "Two functional thioredoxins containing redox-sensensitive vicinal
RL dithiols from the Chlamydomonas outer dynein arm."
RJ J. Biol. Chem. 271:6283-6291 (1996).
CC -1- FUNCTION: MAY BE INVOLVED IN REGULATING THE REDOX STATE OF
CC FUNCTIONALLY IMPORTANT THIOL GROUPS WITHIN DYNEIN.
CC -1- SUBUNIT: CONSISTS OF AT LEAST 3 HEAVY CHAINS (ALPHA, BETA AND
CC GAMMA), 2 INTERMEDIATE CHAINS AND 8 LIGHT CHAINS.
CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U43609; AAB03681.1; -
DR PIR; T08084; T08084.
DR InterPro; IPR006662; ThioRed.
DR InterPro; IPR006663; ThioRedox_dom2.
DR Pfam; PF00085; ThioRed; 1.
DR PROSITE; PS00194; THIOREDOXIN; 1.
KW Microtubules; Dynein; Flagella; Redox-active center;
FT Electron transport.
FT DISULFID 34 37 REDOX-ACTIVE (BY SIMILARITY).
SQ SEQUENCE 129 AA; 14179 MW; 3E3F0B02E8C53E09 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 129;
Best Local Similarity 100.0%; Pred.No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPA 147
DB 119 AAPAAPA 125

RESULT 26
SIB2 ECOLI
ID SIB2_ECOLI STANDARD; PRT; 130 AA.
AC Q467E5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein sirB2.
OS Escherichia coli.
GN SIRB2 OR B1213.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RT SEQUENCE FROM N.A.
RX MEDLINE=95362678; PubMed=7543480;
RA Strohmaier H., Renner P., Renner W., Hoegenauer G.;
RT "Expression of genes kdsA and kdsB involved in 3-deoxy-D-manno-
```

```

octulosonic acid metabolism and biosynthesis of enterobacterial
lipopolysaccharide is growth phase regulated primarily at the
transcriptional level in Escherichia coli K-12."
RJ J. Bacteriol. 177:4488-4500 (1995).
RN [2]
RT SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba T., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RL corresponding to the 12.7-28.0 min region on the linkage map."
RC DNA Res. 3:137-155 (1996).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U18555; AAC43439.1; -
DR EMBL; AE000219; AAC74297.1; -
DR EMBL; D90757; BAA36081.1; -
DR EMBL; D90756; BAA36071.1; -
DR PIR; I83571; I83571.
DR EcoGene; EG14293; sirB2.
DR Pfam; PF04247; SirB; 1.
KW Transmembrane; Complete proteome.
FT TRANSMEM 10 28 POTENTIAL.
FT TRANSMEM 40 59 POTENTIAL.
FT TRANSMEM 74 93 POTENTIAL.
FT TRANSMEM 105 124 POTENTIAL.
SQ SEQUENCE 130 AA; 14639 MW; 58A2D27776F44238 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 130;
Best Local Similarity 100.0%; Pred.No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 724 TKVPLLG 730
DB 124 TKVPLLG 130

RESULT 27
YEDX_ECO57
ID YEDX_ECO57 STANDARD; PRT; 137 AA.
AC Q8VB75;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transhyretin-like protein precursor.
GN YEDX OR Z3062 OR ECS2708.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
```

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX  MEDLINE=21074935; PubMed=11206551;
RA  Perna N.T., Plunkett G. III, Burland V., Bloch C.A., Perna N.T., Burland V.,
RA  Riley M., Mayhew G.F., Glasner J.D., Glasner J.D., Mayhew G.F.,
RA  Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA  Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA  Grobeck E.U., Davis N.W., Lim A., Dimalanta E.T., Potamouzis K.,
RA  Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA  Welch R.A., Blattner F.R.;
RA  "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL  Nature 409:529-533(2001).
[2]
RP  SEQUENCE FROM N.A.
RX  STRAIN=O157:H7 / RMD 050952;
RX  MEDLINE=2115231; PubMed=11258795;
RA  Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA  Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA  Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA  Kihara S., Shiba T., Hattori M., Shinagawa H.;
RA  "Complete genome sequence of enterohemorrhagic Escherichia coli
RT  O157:H7 and genomic comparison with a laboratory strain K-12."
RL  DNA Res. 8:11-22(2001).
CC  -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC  -1- SIMILARITY: BELONGS TO THE TRANSTHYRETIN FAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AE005418; AAG56983.1; -
CC  PIR; C85815; C85815.
CC  PIR; D90967; D90967.
CC  InterPro; IPR000895; Transthyretin.
CC  Pfam; PF00576; Transthyretin; 1.
CC  ProDom; PD003457; Transthyretin; 1.
CC  SMART; SM00095; TR_THY; 1.
CC  PROSITE; PS00768; TRANSTHYRETIN_1; 1.
CC  PROSITE; PS00769; TRANSTHYRETIN_2; 1.
CC  TRANSPORT; Periplasmic; Signal; Complete proteome.
CC  SIGNAL 1 23
CC  FT CHAIN 24 137
CC  SQ SEQUENCE 137 AA; 15537 MW; 0A8D534E6A46B9F8 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 137;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 KQONLES 333
DB 96 KQONLES 102

RESULT 28
YEDX_ECOLI STANDARD; PRT; 137 AA.
AC P76341;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transthyretin-like protein precursor.
GN YEDX OR B1970.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]

RN  SEQUENCE FROM N.A.
RP  STRAIN=K12 / MG1655;
RX  MEDLINE=97426617; PubMed=9278503;
RA  Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA  Riley M., Collado-Vides J., Glasner J.D., Glasner J.D., Mayhew G.F.,
RA  Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA  Mau B., Shao Y.;
RA  "The complete genome sequence of Escherichia coli K-12."
RL  Science 277:123-124(1997).
[2]
RP  PARTIAL SEQUENCE OF N-TERMINUS.
RA  Rudd K.E.;
RA  Unpublished observations (AUG-1999).
CC  -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC  -1- SIMILARITY: BELONGS TO THE TRANSTHYRETIN FAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AE000288; AAC75036.1; -
CC  PIR; F64961; F64961.
CC  RSP; P27731; 1TTP.
CC  EcoGene; EG14046; yedX.
CC  InterPro; IPR000895; Transthyretin.
CC  Pfam; PF00576; Transthyretin; 1.
CC  PRINTS; PR00189; TRANSTHYRETIN.
CC  ProDom; PD003457; Transthyretin; 1.
CC  SMART; SM00095; TR_THY; 1.
CC  PROSITE; PS00768; TRANSTHYRETIN_1; 1.
CC  PROSITE; PS00769; TRANSTHYRETIN_2; 1.
CC  TRANSPORT; Periplasmic; Signal; Complete proteome.
CC  SIGNAL 1 23
CC  FT CHAIN 24 137
CC  SQ SEQUENCE 137 AA; 15460 MW; BS8AS34051DDDC5E CRC64;

Query Match 0.9%; Score 7; DB 1; Length 137;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 KQONLES 333
DB 96 KQONLES 102

RESULT 29
YOR1_CALSR STANDARD; PRT; 140 AA.
AC P40079;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein in xylR 5' region (ORF1) (Fragment).
OS Caldicellulosiruptor sp. (strain Rt8B.4).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=28238;
RN [1]

RN  SEQUENCE FROM N.A.
RP  MEDLINE=97077616; PubMed=8920183;
RA  Divedi P.P., Gibbs M.D., Saul D.J., Bergquist P.L.;
RA  "Cloning, sequencing and overexpression in Escherichia coli of a
RT  xylanase gene, xynA from the thermophilic bacterium Rt8B.4 genus
RT  Caldicellulosiruptor."
RL  Appl. Microbiol. Biotechnol. 45:86-93(1996).
CC  -1- FUNCTION: MAY PLAY A ROLE IN SUGAR TRANSPORT.
CC  -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC  -1- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
CC  SYSTEM PERMEASE FAMILY. MALFG SUBFAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L18965; AAB42041.1; -.
DR PIR; S41785; S41785.
DR InterPro; IPR000515; BPD transp.
DR Pfam; PF00528; BPD transp. 1.
DR PROSITE; PS00402; BPD TRANSP INN MEMBER; 1.
KW Hypothetical protein; Transmembrane; Transport.
FT NONTER 1
FT TRANSMEM 9
FT TRANSMEM 29
FT TRANSMEM 65
FT TRANSMEM 85
FT TRANSMEM 115
FT TRANSMEM 135
SQ SEQUENCE 140 AA; 15761 MW; FAD18780D92692BF CRC64;

Query Match 0.9%; Score 7; DB 1; Length 140;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 300 LIITTAG 306
Db 122 LIITTAG 128
|||||

RESULT 30
BCCP_BACSU STANDARD; PRT; 159 AA.
AC P49786;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Biotin carboxyl carrier protein of acetyl-CoA carboxylase (BCCP).
GN ACCB OR FASE.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96074336; PubMed=7592499;
RA Marini P.E., Li S.U., Gardiol D., Cronan J.E. Jr., de Mendoza D.;
RT "The genes encoding the biotin carboxyl carrier protein and biotin
RT carboxylase subunits of Bacillus subtilis acetyl coenzyme A
RT carboxylase, the first enzyme of fatty acid synthesis.";
RL J. Bacteriol. 177:7003-7006(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=97124195; PubMed=9969508;
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
RA Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
RT the Bacillus subtilis genome containing the skin element and many
RT sporulation genes";
RL Microbiology 142:3103-3111(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Besieres P., Bolotin A., Borchert S.,
RA Borriais R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,

```

```

RA Chim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetle D., Purnell S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Solido B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [4]
RP SEQUENCE OF 1-124 FROM N.A.
RC STRAIN=168 / JH642;
RA Guerout-Fleury A.M., Gonzy-Treboul G., Stragier P.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACETYL COENZYME A
CC CARBOXYLASE COMPLEX; FIRST, BIOTIN CARBOXYLASE CATALYZES THE
CC CARBOXYLATION OF THE CARRIER PROTEIN AND THEN THE TRANSCARBOXYLASE
CC TRANSFERS THE CARBOXYL GROUP TO FORM MALONYL-COA.
CC -!- PATHWAY: Long-chain fatty acid biosynthesis; first step.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U36245; AAB00182.1; -.
DR EMBL; D84432; BAA12568.1; -.
DR EMBL; Z99116; CAB14366.1; -.
DR EMBL; U35252; AAA76728.1; -.
DR PIR; H69580; H69580.
DR HSP; P02905; 3BDO.
DR Subtilist; BG11383; accB.
DR InterPro; IPR001249; ACCoA_biotinCC.
DR InterPro; IPR001882; Biotin_attach.
DR Pfam; PF00364; biotin_lipoyl.
DR Pfam; PF00364; biotin_lipoyl; 1.
DR PRINTS; PR01071; ACCoABiotinCC.
DR TIGRFAMs; TIGR00531; BCCP; 1.
DR PROSITE; PS00186; BIOTIN; 1.
KW Fatty acid biosynthesis; Biotin; Complete proteome.
FT BINDING 123
FT BINDING 123
FT CONFLICT 64
FT CONFLICT 65
FT CONFLICT 126
FT CONFLICT 126
FT CONFLICT 126
FT CONFLICT 126
SQ SEQUENCE 159 AA; 17228 MW; 7A1COA1E2703A07F CRC64;

Query Match 0.9%; Score 7; DB 1; Length 159;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 QQAAPA 223
Db 65 QQAAPA 71
|||||

RESULT 31

```

TH12 HUMAN  
 ID TH12 HUMAN STANDARD; PRT; 166 AA.  
 AC Q99757; Q9UH29;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Thioredoxin, mitochondrial precursor (MT-TRX) (Thioredoxin 2).  
 GN TXN2 OR TRX2.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RC Miranda-Vizuela A., Gustafsson J.-A., Spyrou G.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Lens;  
 RA Ready P.G., Bhuvan D.K., Bhuvan K.C.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=20057165; PubMed=10591208;  
 RA Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,  
 RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,  
 RA Baguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,  
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,  
 RA Burdill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,  
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
 RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,  
 RA Evans K.L., Fey J.K., Fleming K., French L., Garner A.A.,  
 RA Gilbert J.G.R., Goward M.E., Grahham D.V., Griffiths M.N.D., Hall C.,  
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,  
 RA Hunt S.E., Jones M.C., Kerhaw J., Kimberley A.M., King A.,  
 RA Laird G.K., Langford C.P., Leverhulme M.A., Lloyd C., Lloyd D.M.,  
 RA Martyn I.D., Mashregi-Mohammadi M., Matthews L.H., McCann O.T.,  
 RA McClay J., McLaren S., McMurry A.V., Milne S.A., Mortimore B.J.,  
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,  
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,  
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,  
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,  
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,  
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,  
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Mala J.E., Nguyen T., Pan H.,  
 RA Phan S., Qi S., Qian Y., Ray U., Ren Q., Shaull S., Sloan D., Song L.,  
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,  
 RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,  
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,  
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,  
 RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,  
 RA Korfi I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,  
 RA Emanuel B.S., Shaikh T., Kurakshi H., Saitta S., Budarf M.B.,  
 RA McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,  
 RA Kim U.J., Shizuya H., Simon M.I., Dunanski J.P., Peyrard M., Kedra D.,  
 RA Seroussi E., Franssion I., Tapia I., Bruder C.E., O'Brien K.P.,  
 RA Wilkison P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,  
 RA Tilahun Y., Wright H.;  
 RT "The DNA sequence of human chromosome 22.";  
 RL Nature 402:499-495(1999).  
 [4]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=22386257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Frange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S.S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Maman A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Maman A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).  
 CC -!- FUNCTION: POSSES A DITHIOL-REDUCING ACTIVITY.  
 CC -!- SUBCELLULAR LOCATION: Mitochondrial.  
 CC -!- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 ENBL; U78678; AAB41631.1; -  
 DR ENBL; AF278920; AAF98467.1; -  
 DR ENBL; AL022313; CAA18439.1; -  
 DR ENBL; BC013726; AAH13726.1; -  
 DR HSSP; P10599; 1TRV.  
 DR Genew; HGNC:17772; TXN2.  
 DR GO; GO:0000008; F:thioredoxin; TAS.  
 DR InterPro; IPR006662; Thiored.  
 DR InterPro; IPR006663; Thioredox\_dom2.  
 DR InterPro; IPR005746; Thioredoxin.  
 DR Pfam; PF00085; Thiored; 1.  
 DR PRINTS; PR00421; THIOREDIXIN.  
 DR TIGRFAMs; TIGR01068; thioredoxin; 1.  
 DR PROSITE; PS00194; THIOREDIXIN; 1.  
 DR Redox-active center; Electron transport; Mitochondrion;  
 DR Transist peptide.  
 KW TRANSIT 1 59 MITOCHONDRION (BY SIMILARITY).  
 KW CHAIN 60 166 THIOREDIXIN.  
 FT DISULFID 90 93 REDOX-ACTIVE (BY SIMILARITY).  
 FT CONFLICT 30 30 R -> K (IN REF. 1 AND 2).  
 SQ SEQUENCE 166 AA; 18383 MW; C4CA8CDAD485D499 CRC64;  
 Query Match 0.9%; Score 7; DB 1; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 655 GLTVTPN 661 PRT; 167 AA.  
 DB 41 GLTVTPN 47  
 |||||  
 RESULT 32  
 HES5 MOUSE  
 AC P70120;  
 ID HES5 MOUSE STANDARD; PRT; 167 AA.  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Transcription factor HES-5 (Hairy and enhancer of split 5).  
 GN HES5 OR HES-5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SV; TISSUE=Liver;  
RX MEDLINE=95138130; PubMed=7836401;  
RA Takebayashi K., Akazawa C., Nakanishi S., Kageyama R.;  
RT "Structure and promoter analysis of the gene encoding the mouse  
RT helix-loop-helix factor HES-5. Identification of the neural  
RT precursor cell-specific promoter element.";  
RL J. Biol. Chem. 270:1342-1349(1995)  
CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR OF GENES THAT REQUIRE A BHLH  
CC PROTEIN FOR THEIR TRANSCRIPTION.  
CC WITH A CO-REPRESSOR PROTEIN (GROUCHO).  
CC -!- SUBUNIT: TRANSCRIPTION REPRESSION REQUIRES FORMATION OF A COMPLEX  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- DOMAIN: HAS A PARTICULAR TYPE OF BASIC DOMAIN (PRESENCE OF A  
CC HELIX-INTERUPTING PROLINE) THAT BINDS TO THE N-BOX (CACNAG),  
CC RATHER THAN THE CANONICAL E-BOX (CANNTG).  
CC -!- DOMAIN: THE CARBOXYL-TERMINAL WRPW MOTIF IS A TRANSCRIPTIONAL  
CC REPRESSION DOMAIN NECESSARY FOR THE INTERACTION WITH GROUCHO. A  
CC TRANSCRIPTIONAL CO-REPRESSOR RECRUITED TO SPECIFIC TARGET DNA BY  
CC HAIRY-RELATED PROTEINS.  
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
CC TRANSCRIPTION FACTORS. "HAIRY" SUBFAMILY.  
CC -!- SIMILARITY: Contains 1 orange domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D32132; BAA06858.1; -  
DR PIR; A55438;  
DR MGD; MGI:104876; Hes5.  
DR InterPro; IPR001092; HLH\_basic.  
DR InterPro; IPR003650; Orange.  
DR Pfam; PF00010; HLH; 1  
DR SMART; SM00353; HLH; 1.  
DR SMART; SM00511; ORANGE; 1.  
DR PROSITE; PS00038; HLH\_1; FALSE\_NEG.  
DR PROSITE; PS00888; HLH\_2; 1.  
DR Developmental protein; Neurogenesis; Nuclear protein; DNA-binding;  
KW Transcription regulation; Repressor.  
FT DOMAIN 14 31 ARG/LYS-RICH (BASIC).  
FT DNA BIND 17 29 BASIC DOMAIN  
FT DOMAIN 30 73 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
FT DOMAIN 74 79 POLY-ALA.  
FT DOMAIN 122 141 PRO-RICH.  
FT DOMAIN 146 152 POLY-ALA.  
FT DOMAIN 164 167 WRPW MOTIF (REQUIRED FOR ACTIVITY)  
FT (BY SIMILARITY).  
SQ SEQUENCE 167 AA; 18425 MW; CA0E697094F4E760 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 167;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 142 APAAPAK 148  
DB 124 APAAPAK 130  
|||||

RESULT 33  
PER2 HORVU STANDARD; PRT; 170 AA.  
ID PER2 HORVU  
AC Q01548;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Peroxidase 2 (EC 1.11.1.7) (Fragment).  
OS Hordeum vulgare (Barley).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
OC Triticeae; Hordeum.  
OX NCBI\_TaxID=4513;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Carina; TISSUE=Leaf;  
RA Thorald-Christensen H., Brandt J., Cho B.H., Raemussen S.K.,  
RA Gregersen P.L., Smedegaard-Petersen V., Collinge D.B.;  
RT "cDNA cloning and characterization of two barley peroxidase  
RT transcripts induced differentially by the powdery mildew fungus  
RT Erysiphe graminis.";  
RL Physiol. Mol. Plant Pathol. 40:395-409(1992).  
CC -!- FUNCTION: Removal of H(2)O(2), oxidation of toxic reductants,  
CC biosynthesis and degradation of lignin, suberization, auxin  
CC catabolism, response to environmental stresses such as wounding,  
CC pathogen attack and oxidative stress. These functions might be  
CC dependent on each isozyme/isoform in each plant tissue.  
CC -!- FUNCTION: Involved in defense response to powdery mildew fungus.  
CC -!- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.  
CC -!- COFACTOR: Binds 1 protoheme IX, 1 iron(III) ion and 2 calcium  
CC ions.  
CC -!- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. CLASSICAL PLANT  
CC (CLASS III) PEROXIDASE SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X62438; CAA44304.1; -  
DR PIR; S18064; S18064.  
DR HSSP; P22195; 1SCH.  
DR InterPro; IPR002016; Peroxidase.  
DR Pfam; PF00141; Peroxidase; 1.  
DR PRINTS; PR00459; PEROXIDASE.  
DR PROSITE; PS00435; PEROXIDASE\_1; 1.  
DR PROSITE; PS00436; PEROXIDASE\_2; PARTIAL.  
DR PROSITE; PS00873; PEROXIDASE\_4; 1.  
KW Oxidoreductase; Glycoprotein; Peroxidase; Iron; Heme;  
KW Multigene family; Calcium.  
FT NON\_TER 1 1  
FT METAL 25 25 CALCIUM 2 (BY SIMILARITY).  
FT METAL 73 73 CALCIUM 2 (BY SIMILARITY).  
FT METAL 76 76 CALCIUM 2 (BY SIMILARITY).  
FT METAL 81 81 CALCIUM 2 (BY SIMILARITY).  
FT METAL 24 24 IRON (HEME AXIAL LIGAND).  
FT ACT SITE 98 98 HYDROGEN-BOUND (BY SIMILARITY).  
FT DISULFID 31 59 BY SIMILARITY.  
FT CARBOHYD 9 9 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 63 63 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 170 AA; 18882 MW; E64B2C012157732B CRC64;

Query Match 0.9%; Score 7; DB 1; Length 170;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 275 RSLDVAD 281  
DB 10 RSLDVAD 16  
|||||

RESULT 34  
PCR1 SCHPO STANDARD; PRT; 171 AA.  
ID PCR1 SCHPO  
AC Q09926;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Transcription factor pcr1 (transcription factor mts2).  
 GN PCr1 OR MTS2 OR SPAC21E11.03C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 CC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=96140438; PubMed=8552099;  
 RX Watanabe Y., Yamamoto M.;  
 RA "Schizosaccharomyces pombe pcr1+ encodes a CREB/ATF protein involved  
 RT in regulation of gene expression for sexual development.";  
 RL Mol. Cell. Biol. 16:704-711(1996).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP Kon N., Krawchuk M.D., Warren B.G., Smith G.R., Wahls W.P.;  
 RX Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moutle S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitz E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,  
 RA Welljens I., Vansireals E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Reinhardt R., Pohl T.M.,  
 RA Borzym K., Langer I., Beck A., Lehner H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Coffeau A., Cadieu E., Dreano S., Gloux S., Leilaue V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potaashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 RN [4]  
 RN CHARACTERIZATION.  
 RP MEDLINE=95047325; PubMed=7958849;  
 RA Wahls W.P., Smith G.R.;  
 RT "A heteromeric protein that binds to a meiotic homologous  
 RT recombination hot spot: correlation of binding and hot spot  
 RT activity.";  
 RL Genes Dev. 8:1693-1702(1994).  
 CC -!- FUNCTION: INVOLVED IN REGULATION OF GENE EXPRESSION FOR SEXUAL  
 CC DEVELOPMENT. BINDS AND ACTIVATES MEIOTIC RECOMBINATION HOT SPOT  
 CC ADE6-M26.  
 CC -!- SUBUNIT: HETERODIMER OF PCr1/MTS2 AND ATrf1/MTS1.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- SIMILARITY: Belongs to the bzip family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D63667; BAA09818.1; -

DR EMBL: U87870; AAB46991.1; -  
 DR EMBL: Z67999; CAA91968.1; -  
 DR PIR: S62588; S62588.  
 DR TRANSFAC: T01687; -  
 DR GeneDB SPombe: SPAC21E11.03c; -  
 DR InterPro: IPR004827; TF\_bZIP.  
 DR Pfam: PF00170; bZIP; 1.  
 DR SMART: SM00338; BRLZ; 1.  
 DR PROSITE: PS0217; bZIP; 1.  
 DR PROSITE: PS00036; bZIP\_BASIC; 1.  
 DR Transcription regulation; Activator; DNA-binding; Nuclear protein;  
 KW Meiosis.  
 FT DNA BIND 12 32 BASIC MOTIF.  
 FT DOMAIN 42 66 LEUCINE-ZIPPER.  
 SQ SEQUENCE 171 AA; 19348 MW; 9922FDDDFE150BDE CRC64;  
 Query Match 0.9%; Score 7; DB 1; Length 171;  
 Best Local Similarity 100.0%; Pred. No. 50;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 179 ESVVSVS 185  
 Db 116 ESVVSVS 122  
 |||||  
 APT\_SVNY3 STANDARD; PRT; 172 AA.  
 ID\_APT\_SVNY3  
 AC P73935;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Adenine phosphoribosyltransferase (EC 2.4.2.7) (APRT).  
 GN APT OR SLL1430.  
 DE Synecocystis sp. (strain PCC 6803).  
 OS Bacteria; Cyanobacteria; Chroococcales; Synecocystis.  
 OX NCBI\_TaxID=1148;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=97061201; PubMed=8905231;  
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hirosewa M., Sugura M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
 RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M.,  
 RA Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synecocystis sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions.";  
 RL DNA Res. 3:109-136(1996).  
 CC -!- FUNCTION: Catalyzes a salvage reaction resulting in the formation  
 CC of AMP, that is energetically less costly than de novo synthesis.  
 CC -!- CATALYTIC ACTIVITY: AMP + diphosphate = adenine + 5-phospho-  
 CC alpha-D-ribose 1-diphosphate.  
 CC -!- PATHWAY: Purine salvage.  
 CC -!- SUBUNIT: Homodimer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: BELONGS TO THE PURINE/PYRIMIDINE  
 CC PHOSPHORIBOSYLTRANSFERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D90911; BAA18001.1; -  
 DR PIR: S75440; S75440.  
 DR HAMAP: MF 00004; -; 1.  
 DR InterPro: IPR005764; Ade\_phospho\_trans.  
 DR InterPro: IPR002375; Pr/By\_tp\_transf.  
 DR InterPro: IPR000836; PRtransferase.

```

DR Pfam; PF00156; Priboosyltran; 1.
DR TIGRFAMs; TIGR01090; apt; 1.
DR PROSITE; PS00103; PUR_PVR_PR_TRANSFER; 1.
KW Transferrase; Glycosyltransferase; Purine salvage; Complete proteome.
SQ SEQUENCE 172 AA; 18997 MW; D57E94BCEFAE457 CRC64;

Query Match          0.9%; Score 7; DB 1; Length 172;
Best Local Similarity 100.0%; Pred.No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 IIELAAL 245
    |||||
Db 149 IIELAAL 155

RESULT 36
YMF3_CAEEL STANDARD; PRT; 175 AA.
AC P34463;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein F55H2.3 in chromosome III.
GN F55H2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hallier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden K.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; J03807; AAA30055.1; -
DR PIR; A32137; A32137.
DR HSP; P02259; IHST.
DR InterPro; IPR005818; Histone H1/H5.
DR InterPro; IPR003216; Linker histone N.
DR Pfam; PF00538; Linker histone; 1.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
DR KEGG; K01001; Histone H1; 1.
RW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ SEQUENCE 185 AA; 19297 MW; DA6F3110F9F946E6 CRC64;

Query Match          0.9%; Score 7; DB 1; Length 185;
Best Local Similarity 100.0%; Pred.No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPA 147
    |||||
Db 6 AAPAAPA 12

RESULT 38
RBS1_CHIRE STANDARD; PRT; 185 AA.
AC P00873;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ribulose biphosphate carboxylase small chain 1, chloroplast precursor
DE (EC 4.1.1.39) (Rubisco small subunit 1).
GN RECS-1.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87141171; PubMed=3820291;
RA Goldschmidt-Clermont M., Rahire M.;
RT "Sequence, evolution and differential expression of the two genes
RT encoding variant small subunits of ribulose biphosphate
RT carboxylase/oxygenase in Chlamydomonas reinhardtii.";
RL J. Mol. Biol. 191:421-432(1986).
RN [2]
RP SEQUENCE OF 1-57.

```

```

DR Pfam; PF00156; Priboosyltran; 1.
DR TIGRFAMs; TIGR01090; apt; 1.
DR PROSITE; PS00103; PUR_PVR_PR_TRANSFER; 1.
KW Transferrase; Glycosyltransferase; Purine salvage; Complete proteome.
SQ SEQUENCE 172 AA; 18997 MW; D57E94BCEFAE457 CRC64;

Query Match          0.9%; Score 7; DB 1; Length 172;
Best Local Similarity 100.0%; Pred.No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 IIELAAL 245
    |||||
Db 149 IIELAAL 155

RESULT 36
YMF3_CAEEL STANDARD; PRT; 175 AA.
AC P34463;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein F55H2.3 in chromosome III.
GN F55H2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hallier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden K.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; J03807; AAA30055.1; -
DR PIR; A32137; A32137.
DR HSP; P02259; IHST.
DR InterPro; IPR005818; Histone H1/H5.
DR InterPro; IPR003216; Linker histone N.
DR Pfam; PF00538; Linker histone; 1.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
DR KEGG; K01001; Histone H1; 1.
RW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ SEQUENCE 185 AA; 19297 MW; DA6F3110F9F946E6 CRC64;

Query Match          0.9%; Score 7; DB 1; Length 175;
Best Local Similarity 100.0%; Pred.No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 527 DGFGRDL 533
    |||||
Db 3 DGFGRDL 9

RESULT 37
H1D_STRPU STANDARD; PRT; 185 AA.
ID_H1D_STRPU

```

```

RX MEDLINE=80094642; PubMed=521455;
RA Schmidt G.W., Devillers-Thiery A., Desruisseaux H., Biobel G.,
RA Chua N.-H.;
RT "NH2-terminal amino acid sequences of precursor and mature forms of
RT the ribulose-1,5-bisphosphate carboxylase small subunit from
RT Chlamydomonas reinhardtii."
RL J. Cell Biol. 83:615-622(1979).
CC -!- FUNCTION: RuBisCO catalyzes two reactions: the carboxylation of D-
CC ribulose 1,5-bisphosphate, the primary event in photosynthetic
CC carbon dioxide fixation, as well as the oxidative fragmentation of
CC the pentose substrate in the photorespiration process. Both
CC reactions occur simultaneously and in competition at the same
CC active site.
CC -!- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-
CC phospho-D-glycerate.
CC -!- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) =
CC 3-phospho-D-glycerate + 2-phosphoglycolate.
CC -!- SUBUNIT: 8 large chains + 8 small chains.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X04471; CAA28159.1; ALT_INIT.
CC PIR; A25785; RKKMS1.
CC DR PDB; 1GX8; 24-OCT-01.
CC DR InterPro; IPR000894; RuBisCO small.
CC DR Pfam; PF00101; RuBisCO small; 1.
CC DR PRINTS; PR00152; RUBISCO SMALL.
CC DR ProDom; PD000290; RuBisCO small; 1.
CC KW Photosynthesis; Carbon dioxide fixation; Photorespiration; Lyase;
CC KW Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide;
CC KW Multigene family; 3D-structure.
CC FT TRANSIT 1 45 CHLOROPLAST.
CC FT CHAIN 46 185 RUBULOSE BISPHOSPHATE CARBOXYLASE SMALL
CC FT CHAIN 1.
CC FT CONFLICT 2 2 MISSING (IN REF. 2).
CC FT CONFLICT 42 42 Q -> E (IN REF. 2).
CC FT CONFLICT 45 45 Q -> D (IN REF. 2).
CC SQ SEQUENCE 185 AA; 20620 MW; B4114FD98E807F16 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 185;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 PAVKAAP 143
Db 30 PAVKAAP 36
|||||
RESULT 39
RBS2_CHLRE STANDARD; PRT; 185 AA.
AC P08475;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ribulose biphosphate carboxylase small chain 2, chloroplast precursor
DE (EC 4.1.1.39) (RuBisCO small subunit 2).
GN RECS-2.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RP MEDLINE=87141171; PubMed=3920291;

```

```

RA Goldschmidt-Clermont M., Rahire M.;
RT "Sequence, evolution and differential expression of the two genes
RT encoding variant small subunits of ribulose biphosphate
RT carboxylase/oxygenase in Chlamydomonas reinhardtii."
RL J. Mol. Biol. 191:421-432(1986).
CC -!- FUNCTION: RuBisCO catalyzes two reactions: the carboxylation of D-
CC ribulose 1,5-bisphosphate, the primary event in photosynthetic
CC carbon dioxide fixation, as well as the oxidative fragmentation of
CC the pentose substrate in the photorespiration process. Both
CC reactions occur simultaneously and in competition at the same
CC active site.
CC -!- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-
CC phospho-D-glycerate.
CC -!- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) =
CC 3-phospho-D-glycerate + 2-phosphoglycolate.
CC -!- SUBUNIT: 8 large chains + 8 small chains.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X04472; CAA28160.1; -.
CC PIR; B25785; RKXMS2.
CC DR PDB; 1IR2; 20-MAR-02.
CC DR InterPro; IPR000894; RuBisCO small.
CC DR Pfam; PF00101; RuBisCO small; 1.
CC DR PRINTS; PR00152; RUBISCO SMALL.
CC DR ProDom; PD000290; RuBisCO small; 1.
CC KW Photosynthesis; Carbon dioxide fixation; Photorespiration; Lyase;
CC KW Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide;
CC KW Multigene family; 3D-structure.
CC FT TRANSIT 1 45 CHLOROPLAST.
CC FT CHAIN 46 185 RUBULOSE BISPHOSPHATE CARBOXYLASE SMALL
CC FT CHAIN 2.
CC SQ SEQUENCE 185 AA; 20647 MW; E19A3627EF484F50 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 185;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 PAVKAAP 143
Db 30 PAVKAAP 36
|||||
RESULT 40
RL9_ARATH STANDARD; PRT; 194 AA.
AC P49209; Q9LDX2;
DT 01-FEB-1996 (Rel. 33, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L9.
GN RPL9B OR AT1G33120 OR T9L6.2 AND (RPL9C OR AT1G33140 OR T9L6.5).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia;
RA Grellet F., Cooke R., Laudie M., Raynal M., Delseny M.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia;

```

```

RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Cressey T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maity R., Marziani A.,
RA Millscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:816-820(2000).
RN [3]
RN SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Chao Q., Chen H., Kim C., Ecker J.;
RT "Full length cDNA sequences of Arabidopsis thaliana.";
RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
RT SSP consortium (Salk/Stanford/PGEC).";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [5]
RN SEQUENCE OF 1-42 AND 150-195 FROM N.A.
RC STRAIN=cv. Columbia;
RA Raynal M., Greillet F., Laudie M., Meyer Y., Cooke R., Delseny M.;
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE L6P FAMILY OF RIBOSOMAL PROTEINS.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
CC in positions 134, 139 and 141.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X91958; CAA63024.1; ALT_FRAME.
DR EMBL; AC021045; AAF97348.1; -.
DR EMBL; AC021045; AAF97345.1; -.
DR EMBL; AF324688; AAG40039.1; -.
DR EMBL; AF326873; AAG41455.1; -.
DR EMBL; AF339594; AAK00376.1; -.
DR EMBL; AY058051; AAL24159.1; -.
DR EMBL; AF375419; AAK53003.1; -.
DR EMBL; AY039593; AAK62648.1; -.
DR EMBL; AY054156; AAL06817.1; -.
DR EMBL; AY072446; AAL62438.1; -.
DR EMBL; Z17727; CAA79045.1; -.
DR EMBL; Z17728; CAA79046.1; -.
DR PIR; F86455; F86455.
DR InterPro; IPR000702; Ribosomal_L6.
DR InterPro; IPR002359; Ribosomal_L6_2.
DR Pfam; PF00347; Ribosomal_L6; 2.
DR PROSITE; PS00700; RIBOSOMAL_L6_2; 1.
KW Ribosomal protein.
FT CONFLICT 12 12 I -> R (IN REF. 5; CAA79045).
FT CONFLICT 71 71 A -> V (IN REF. 1).
FT CONFLICT 78 78 S -> R (IN REF. 1).
FT CONFLICT 106 106 A -> V (IN REF. 1).
FT CONFLICT 113 114 KS -> SL (IN REF. 1).

```

```

SQ SEQUENCE 194 AA; 22017 MW; 451874EAE5838ECD CRC64;
Query Match 0.9%; Score 7; DB 1; Length 194;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 540 TGKKKLLK 546
|||
DB 53 TGKKKLLK 59

```

Search completed: December 9, 2003, 10:34:37  
Job time : 32 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 10:32:14 ; Search time 42 Seconds  
(without alignments)

4724.820 Million cell updates/sec

Title: US-09-701-271A-2

Perfect score: 769

Sequence: 1 MNTKLTAKIISGLEFVATAAFQ.....ELLIFITPRIMTAGNSLRY 769

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 256052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

SPTREML 23:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mbc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID     | Description        |
|------------|-------|-------------|--------|-----------|--------------------|
| 1          | 548   | 71.3        | 761    | 16 Q9JYW4 | Q9JYW4 neisseria m |
| 2          | 150   | 19.5        | 720    | 2 Q50972  | Q50972 neisseria g |
| 3          | 150   | 19.5        | 766    | 2 Q9ZHP3  | Q9ZHP3 neisseria m |
| 4          | 11    | 1.4         | 689    | 16 Q9ZN56 | Q9ZN56 rhizobium m |
| 5          | 11    | 1.4         | 714    | 16 Q8XV60 | Q8XV60 raietonia s |
| 6          | 11    | 1.4         | 723    | 2 Q9AM59  | Q9AM59 acinetobact |
| 7          | 10    | 1.3         | 585    | 2 Q32566  | Q32566 escherichia |
| 8          | 10    | 1.3         | 596    | 2 Q47423  | Q47423 escherichia |
| 9          | 10    | 1.3         | 642    | 2 Q9ZGU0  | Q9ZGU0 escherichia |
| 10         | 10    | 1.3         | 673    | 16 Q8DDT0 | Q8DDT0 vibrio vuln |
| 11         | 10    | 1.3         | 684    | 16 Q8EK21 | Q8EK21 shewanella  |
| 12         | 10    | 1.3         | 750    | 2 Q9ZP86  | Q9ZP86 burkholderi |
| 13         | 10    | 1.3         | 901    | 2 Q9ZP81  | Q9ZP81 myxococcus  |
| 14         | 10    | 1.3         | 908    | 5 Q96516  | Q96516 caenorhabdi |
| 15         | 10    | 1.3         | 909    | 5 Q76618  | Q76618 caenorhabdi |
| 16         | 9     | 1.2         | 214    | 16 Q8RCX3 | Q8RCX3 thermoanaer |

|    |   |     |      |    |        |
|----|---|-----|------|----|--------|
| 17 | 9 | 1.2 | 248  | 5  | O15562 |
| 18 | 9 | 1.2 | 379  | 16 | Q985G3 |
| 19 | 9 | 1.2 | 460  | 2  | Q9JRS7 |
| 20 | 9 | 1.2 | 460  | 2  | Q9X6J1 |
| 21 | 9 | 1.2 | 460  | 2  | Q8GD00 |
| 22 | 9 | 1.2 | 470  | 16 | Q9CMH4 |
| 23 | 9 | 1.2 | 512  | 2  | Q9AJ21 |
| 24 | 9 | 1.2 | 607  | 2  | Q9JKT1 |
| 25 | 9 | 1.2 | 607  | 16 | O56974 |
| 26 | 9 | 1.2 | 616  | 2  | Q8VRN0 |
| 27 | 9 | 1.2 | 616  | 2  | Q8VPC8 |
| 28 | 9 | 1.2 | 649  | 2  | Q9ZFY0 |
| 29 | 9 | 1.2 | 654  | 16 | Q8CVN5 |
| 30 | 9 | 1.2 | 658  | 2  | Q8GBE6 |
| 31 | 9 | 1.2 | 705  | 16 | O66850 |
| 32 | 9 | 1.2 | 783  | 2  | Q9F1Q1 |
| 33 | 8 | 1.0 | 142  | 5  | Q93995 |
| 34 | 8 | 1.0 | 142  | 5  | Q94000 |
| 35 | 8 | 1.0 | 149  | 10 | Q8LN85 |
| 36 | 8 | 1.0 | 159  | 5  | Q962S9 |
| 37 | 8 | 1.0 | 184  | 10 | Q94GX7 |
| 38 | 8 | 1.0 | 215  | 16 | Q98BN4 |
| 39 | 8 | 1.0 | 222  | 16 | Q8XR79 |
| 40 | 8 | 1.0 | 231  | 16 | Q9JXV0 |
| 41 | 8 | 1.0 | 231  | 16 | Q9UVZ0 |
| 42 | 8 | 1.0 | 268  | 16 | Q8FGE1 |
| 43 | 8 | 1.0 | 306  | 16 | Q8YDP3 |
| 44 | 8 | 1.0 | 311  | 16 | Q9PKN3 |
| 45 | 8 | 1.0 | 339  | 10 | Q9FQ77 |
| 46 | 8 | 1.0 | 339  | 16 | Q8FU06 |
| 47 | 8 | 1.0 | 355  | 2  | Q46023 |
| 48 | 8 | 1.0 | 356  | 10 | Q8VW11 |
| 49 | 8 | 1.0 | 360  | 10 | Q9LU24 |
| 50 | 8 | 1.0 | 401  | 10 | Q43148 |
| 51 | 8 | 1.0 | 412  | 16 | Q8ZLK2 |
| 52 | 8 | 1.0 | 412  | 16 | Q8Z206 |
| 53 | 8 | 1.0 | 412  | 16 | Q8X818 |
| 54 | 8 | 1.0 | 412  | 16 | Q8CVW9 |
| 55 | 8 | 1.0 | 444  | 11 | Q9CT46 |
| 56 | 8 | 1.0 | 469  | 16 | Q9A6E8 |
| 57 | 8 | 1.0 | 476  | 5  | Q93996 |
| 58 | 8 | 1.0 | 478  | 5  | Q25684 |
| 59 | 8 | 1.0 | 482  | 5  | Q93998 |
| 60 | 8 | 1.0 | 497  | 16 | Q8Z6L1 |
| 61 | 8 | 1.0 | 497  | 16 | P74864 |
| 62 | 8 | 1.0 | 550  | 11 | Q8BQ88 |
| 63 | 8 | 1.0 | 560  | 16 | Q8E927 |
| 64 | 8 | 1.0 | 591  | 2  | Q52291 |
| 65 | 8 | 1.0 | 592  | 16 | Q8PGE6 |
| 66 | 8 | 1.0 | 617  | 16 | Q92W30 |
| 67 | 8 | 1.0 | 623  | 16 | Q98P35 |
| 68 | 8 | 1.0 | 633  | 16 | Q8PH78 |
| 69 | 8 | 1.0 | 635  | 16 | Q8D7V6 |
| 70 | 8 | 1.0 | 637  | 16 | Q9PGC9 |
| 71 | 8 | 1.0 | 648  | 16 | Q8P5V7 |
| 72 | 8 | 1.0 | 691  | 2  | P94767 |
| 73 | 8 | 1.0 | 706  | 16 | Q9FBV6 |
| 74 | 8 | 1.0 | 707  | 11 | Q99K50 |
| 75 | 8 | 1.0 | 707  | 11 | Q8CE30 |
| 76 | 8 | 1.0 | 707  | 11 | Q8CD23 |
| 77 | 8 | 1.0 | 713  | 2  | O52657 |
| 78 | 8 | 1.0 | 737  | 2  | Q44076 |
| 79 | 8 | 1.0 | 776  | 16 | P912M7 |
| 80 | 8 | 1.0 | 794  | 2  | P72249 |
| 81 | 8 | 1.0 | 819  | 16 | Q9L2E6 |
| 82 | 8 | 1.0 | 980  | 16 | Q8Y2Z9 |
| 83 | 8 | 1.0 | 1183 | 5  | Q94447 |
| 84 | 8 | 1.0 | 1394 | 2  | Q9AIU3 |
| 85 | 8 | 1.0 | 1398 | 5  | Q91048 |
| 86 | 8 | 1.0 | 1766 | 5  | Q25668 |
| 87 | 8 | 1.0 | 1785 | 5  | Q25685 |
| 88 | 8 | 1.0 | 2066 | 9  | Q8LTH9 |
| 89 | 8 | 1.0 | 2204 | 16 | O07231 |

|        |              |
|--------|--------------|
| O15562 | nosema locu  |
| Q985G3 | rhizobium l  |
| Q9JRS7 | actinobacil  |
| Q9X6J1 | actinobacil  |
| Q8GD00 | actinobacil  |
| Q9CMH4 | pasteurella  |
| Q9AJ21 | escherichia  |
| Q9JKT1 | yersinia en  |
| O56974 | yersinia pe  |
| Q8VRN0 | escherichia  |
| Q8VPC8 | escherichia  |
| Q9ZFY0 | pseudomonas  |
| Q8CVN5 | escherichia  |
| Q8GBE6 | yersinia en  |
| O66850 | aquifex aeo  |
| Q9F1Q1 | burkholderi  |
| Q93995 | plasmidium   |
| Q94000 | plasmidium   |
| Q8LN85 | oryza sativ  |
| Q962S9 | spodoptera   |
| Q94GX7 | oryza sativ  |
| Q98BN4 | rhizobium l  |
| Q8XR79 | raistonia s  |
| Q9JXV0 | neisseria m  |
| Q9UVZ0 | neisseria m  |
| Q8FGE1 | escherichia  |
| Q8YDP3 | brucella me  |
| Q9PKN3 | chlamydia m  |
| Q9FQ77 | zea mays (m  |
| Q8FU06 | brucella su  |
| Q46023 | corynebacte  |
| Q8VW11 | lotus japon  |
| Q9LU24 | arabidopsis  |
| Q43148 | selaginella  |
| Q8ZLK2 | salmonella   |
| Q8Z206 | salmonella   |
| Q8X818 | escherichia  |
| Q8CVW9 | escherichia  |
| Q9CT46 | mus musculu  |
| Q9A6E8 | caulobacter  |
| Q93996 | plasmidium   |
| Q25684 | plasmidium   |
| Q93998 | plasmidium   |
| Q8Z6L1 | salmonella   |
| P74864 | salmonella   |
| Q8BQ88 | mus musculu  |
| Q8E927 | shewanella   |
| Q52291 | pseudomonas  |
| Q8PGE6 | xanthomonas  |
| Q92W30 | rhizobium m  |
| Q98P35 | rhizobium l  |
| Q8PH78 | xanthomonas  |
| Q8D7V6 | vibrio vuln  |
| Q9PGC9 | xyliella fas |
| Q8P5V7 | xanthomonas  |
| P94767 | erwinia chr  |
| Q9FBV6 | streptomyce  |
| Q99K50 | mus musculu  |
| Q8CE30 | mus musculu  |
| Q8CD23 | mus musculu  |
| O52657 | pseudomonas  |
| Q44076 | aeromonas h  |
| Q912M7 | pseudomonas  |
| P72249 | rhodobacter  |
| Q9L2E6 | streptomyce  |
| Q8Y2Z9 | anabaena sp  |
| Q94447 | calliphora   |
| Q9AIU3 | anaplasma p  |
| Q91048 | caenorhabdi  |
| Q25668 | plasmidium   |
| Q25685 | plasmidium   |
| Q8LTH9 | staphylococ  |
| O07231 | mycobacteri  |

|     |     |      |    |        |                     |     |   |     |     |    |        |                    |
|-----|-----|------|----|--------|---------------------|-----|---|-----|-----|----|--------|--------------------|
| 90  | 1.0 | 3186 | 16 | Q8VRN2 | Q8vkn2 mycobacteri  | 163 | 7 | 0.9 | 163 | 5  | Q9W2E2 | Q9w2e2 drosophila  |
| 91  | 7   | 52   | 5  | Q3TX00 | Q3txq0 caenorhabdi  | 164 | 7 | 0.9 | 167 | 16 | Q8CMV3 | Q8cmv3 pasteurella |
| 92  | 7   | 55   | 16 | Q3JVR0 | Q3jvr0 neisseria m  | 165 | 7 | 0.9 | 167 | 5  | Q8SUQ2 | Q8suq2 encyphalito |
| 93  | 7   | 64   | 15 | Q3JTV5 | Q3jtv5 neisseria m  | 166 | 7 | 0.9 | 167 | 11 | Q8BV11 | Q8bv11 mus musculu |
| 94  | 7   | 72   | 13 | Q8UUC2 | Q8uuc2 oncorhynch   | 167 | 7 | 0.9 | 167 | 16 | Q8P4V4 | Q8p4v4 xanthomonas |
| 95  | 7   | 72   | 16 | Q8FIK3 | Q8fik3 escherichia  | 168 | 7 | 0.9 | 168 | 16 | Q8FSM7 | Q8fsm7 xanthomonas |
| 96  | 7   | 77   | 12 | Q9JGX6 | Q9jgx6 tt virus. o  | 169 | 7 | 0.9 | 169 | 16 | Q8EWA8 | Q8ewa8 mycoplasma  |
| 97  | 7   | 77   | 12 | Q9JGX6 | Q9jgx6 tt virus. o  | 170 | 7 | 0.9 | 171 | 5  | Q8INI6 | Q8ini6 drosophila  |
| 98  | 7   | 78   | 16 | Q9I3C9 | Q9i3c9 pseudomonas  | 171 | 7 | 0.9 | 171 | 13 | Q90W40 | Q90w40 scomber jap |
| 99  | 7   | 80   | 13 | Q9PMW4 | Q9pmw4 gasterosteus | 172 | 7 | 0.9 | 171 | 16 | Q8U7Q3 | Q8u7q3 agrobacteri |
| 100 | 7   | 84   | 2  | Q49085 | Q49085 mycoplasma   | 173 | 7 | 0.9 | 172 | 10 | Q8L9I4 | Q8l9i4 arabidopsis |
| 101 | 7   | 93   | 10 | Q8L3Y9 | Q8l3y9 cryza sativ  | 174 | 7 | 0.9 | 172 | 10 | Q8FMQ8 | Q8fmq8 arabidopsis |
| 102 | 7   | 94   | 16 | Q8NN89 | Q8nn89 corynebacte  | 175 | 7 | 0.9 | 172 | 13 | Q8IBZ3 | Q8ibz3 sardineps m |
| 103 | 7   | 100  | 13 | Q8QGB6 | Q8qgb6 oncorhynch   | 176 | 7 | 0.9 | 174 | 10 | Q40159 | Q40159 lycopersico |
| 104 | 7   | 100  | 13 | Q8QGB7 | Q8qgb7 oncorhynch   | 177 | 7 | 0.9 | 174 | 12 | Q994E3 | Q994e3 porcine ade |
| 105 | 7   | 104  | 17 | Q9Y8Y5 | Q9y8y5 aeropyrum p  | 178 | 7 | 0.9 | 175 | 5  | Q8I191 | Q8i191 drosophila  |
| 106 | 7   | 105  | 16 | Q9JRI1 | Q9jri1 neisseria m  | 179 | 7 | 0.9 | 176 | 2  | Q8GSL2 | Q8gsl2 methylobact |
| 107 | 7   | 106  | 9  | Q9G026 | Q9g026 bacterioph   | 180 | 7 | 0.9 | 177 | 16 | Q8XFM9 | Q8xfm9 xanthomonas |
| 108 | 7   | 111  | 16 | Q92X40 | Q92x40 rhizobium m  | 181 | 7 | 0.9 | 179 | 16 | Q8P4E5 | Q8p4e5 xanthomonas |
| 109 | 7   | 112  | 16 | Q8R7V9 | Q8r7v9 thermoaer    | 182 | 7 | 0.9 | 181 | 2  | Q9EUL0 | Q9eul0 salmonella  |
| 110 | 7   | 116  | 2  | Q8Z2I4 | Q8z2i4 pseudomonas  | 183 | 7 | 0.9 | 182 | 4  | Q8NZ79 | Q8nz79 homo sapien |
| 111 | 7   | 116  | 16 | Q8XP67 | Q8xp67 clostridium  | 184 | 7 | 0.9 | 182 | 5  | Q8T965 | Q8t965 drosophila  |
| 112 | 7   | 118  | 10 | Q944W0 | Q944w0 phytophthor  | 185 | 7 | 0.9 | 184 | 2  | Q9S1D1 | Q9s1d1 zymomonas m |
| 113 | 7   | 119  | 11 | Q8CB11 | Q8cb11 mus musculu  | 186 | 7 | 0.9 | 188 | 2  | P70872 | P70872 bacillus ce |
| 114 | 7   | 120  | 10 | Q94JC3 | Q94jc3 cryza sativ  | 187 | 7 | 0.9 | 188 | 12 | Q89519 | Q89519 dulcamara m |
| 115 | 7   | 124  | 4  | Q96SP8 | Q96sf8 homo sapien  | 188 | 7 | 0.9 | 189 | 16 | Q8KED3 | Q8ked3 bacillus ha |
| 116 | 7   | 129  | 16 | Q8YHP4 | Q8yhp4 brucella me  | 189 | 7 | 0.9 | 193 | 11 | Q99ML1 | Q99ml1 mus musculu |
| 117 | 7   | 130  | 16 | Q8XDP0 | Q8xdp0 escherichia  | 190 | 7 | 0.9 | 194 | 2  | Q9KWA2 | Q9kwa2 agrobacteri |
| 118 | 7   | 130  | 16 | Q8FI01 | Q8fi01 escherichia  | 191 | 7 | 0.9 | 194 | 10 | Q8LCC2 | Q8lcc2 arabidopsis |
| 119 | 7   | 137  | 10 | Q8LKJ5 | Q8lkj5 capsicum an  | 192 | 7 | 0.9 | 194 | 15 | Q9DVC6 | Q9dvc6 human immun |
| 120 | 7   | 137  | 12 | Q8JXQ2 | Q8jxq2 citrus leaf  | 193 | 7 | 0.9 | 195 | 11 | Q9DB00 | Q9db00 mus musculu |
| 121 | 7   | 137  | 12 | Q8JXP8 | Q8jxp8 citrus leaf  | 194 | 7 | 0.9 | 196 | 16 | Q8F189 | Q8f189 leptospira  |
| 122 | 7   | 137  | 12 | Q8JXP2 | Q8jxp2 citrus leaf  | 195 | 7 | 0.9 | 197 | 16 | Q99RB7 | Q99rb7 staphylococ |
| 123 | 7   | 137  | 12 | Q8JXP4 | Q8jxp4 citrus leaf  | 196 | 7 | 0.9 | 198 | 4  | Q8NAN9 | Q8nan9 homo sapien |
| 124 | 7   | 137  | 12 | Q8JXN9 | Q8jxn9 citrus leaf  | 197 | 7 | 0.9 | 198 | 5  | Q9VPW6 | Q9vpw6 drosophila  |
| 125 | 7   | 137  | 12 | Q8JXP5 | Q8jxp5 citrus leaf  | 198 | 7 | 0.9 | 200 | 5  | P82166 | P82166 locusta mig |
| 126 | 7   | 137  | 12 | Q8JXP7 | Q8jxp7 citrus leaf  | 199 | 7 | 0.9 | 201 | 16 | Q8ZE22 | Q8ze22 yersinia pe |
| 127 | 7   | 137  | 12 | Q8JXP9 | Q8jxp9 citrus leaf  | 200 | 7 | 0.9 | 202 | 11 | Q9DAC1 | Q9dac1 mus musculu |
| 128 | 7   | 137  | 12 | Q8JXQ0 | Q8jxq0 citrus leaf  | 201 | 7 | 0.9 | 202 | 16 | Q8DCL3 | Q8dcl3 yersinia pe |
| 129 | 7   | 137  | 12 | Q8JXQ1 | Q8jxq1 citrus leaf  | 202 | 7 | 0.9 | 202 | 17 | Q9P683 | Q9p683 archaeglob  |
| 130 | 7   | 137  | 12 | Q8JXP0 | Q8jxp0 citrus leaf  | 203 | 7 | 0.9 | 202 | 17 | Q8P2P9 | Q8p2p9 methanosarc |
| 131 | 7   | 137  | 12 | Q8JXP1 | Q8jxp1 citrus leaf  | 204 | 7 | 0.9 | 204 | 10 | Q94D57 | Q94d57 oryza sativ |
| 132 | 7   | 137  | 16 | Q8CVY5 | Q8cvy5 escherichia  | 205 | 7 | 0.9 | 205 | 5  | Q95Xf5 | Q95xf5 caenorhabdi |
| 133 | 7   | 138  | 16 | Q8K8B1 | Q8k8b1 bacillus ha  | 206 | 7 | 0.9 | 205 | 16 | Q8NU62 | Q8nu62 corynebacte |
| 134 | 7   | 138  | 16 | Q8X7K6 | Q8x7k6 escherichia  | 207 | 7 | 0.9 | 208 | 10 | Q9AWJ1 | Q9awj1 cryza sativ |
| 135 | 7   | 139  | 2  | Q9X5U4 | Q9x5u4 streptomyce  | 208 | 7 | 0.9 | 208 | 16 | Q8AGJ6 | Q8agj6 rhizobium l |
| 136 | 7   | 139  | 16 | Q8Y2D2 | Q8y2d2 ralstonia s  | 209 | 7 | 0.9 | 210 | 5  | Q8I4S1 | Q8i4s1 plasmodium  |
| 137 | 7   | 140  | 16 | Q8F9K0 | Q8f9k0 leptospira   | 210 | 7 | 0.9 | 210 | 15 | Q8AQT8 | Q8aqt8 human immun |
| 138 | 7   | 140  | 17 | Q9YAO9 | Q9yao9 aeropyrum p  | 211 | 7 | 0.9 | 212 | 5  | Q9XWP7 | Q9xwp7 caenorhabdi |
| 139 | 7   | 143  | 2  | Q8VR45 | Q8vr45 escherichia  | 212 | 7 | 0.9 | 212 | 16 | Q9A7I2 | Q9a7i2 caulobacter |
| 140 | 7   | 144  | 5  | Q00878 | Q00878 plasmodium   | 213 | 7 | 0.9 | 213 | 15 | Q8A991 | Q8a991 chlorobium  |
| 141 | 7   | 145  | 2  | Q9ZAA7 | Q9zaa7 acidaminoco  | 214 | 7 | 0.9 | 213 | 15 | Q8ALC3 | Q8alc3 human immun |
| 142 | 7   | 145  | 12 | Q8QN94 | Q8qn94 ectocarpus   | 215 | 7 | 0.9 | 215 | 5  | Q25256 | Q25256 litomosoid  |
| 143 | 7   | 146  | 16 | Q9K3I4 | Q9k3i4 streptomyce  | 216 | 7 | 0.9 | 215 | 16 | Q8XRT9 | Q8xrt9 ralstonia s |
| 144 | 7   | 147  | 16 | Q92KM6 | Q92km6 rhizobium m  | 217 | 7 | 0.9 | 215 | 16 | Q8PKE2 | Q8pke2 xanthomonas |
| 145 | 7   | 147  | 16 | Q8PAF9 | Q8paf9 xanthomonas  | 218 | 7 | 0.9 | 215 | 16 | Q8E2J2 | Q8e2j2 leptospira  |
| 146 | 7   | 148  | 16 | Q8G0F8 | Q8g0f8 brucella su  | 219 | 7 | 0.9 | 216 | 3  | Q9C426 | Q9c426 ophiostoma  |
| 147 | 7   | 149  | 10 | Q9C8E5 | Q9c8e5 arabidopsis  | 220 | 7 | 0.9 | 216 | 11 | Q9UKY9 | Q9uky9 mus musculu |
| 148 | 7   | 149  | 10 | Q94F02 | Q94fq2 arabidopsis  | 221 | 7 | 0.9 | 216 | 16 | Q67023 | Q67023 aquifex aeo |
| 149 | 7   | 149  | 16 | Q98IA8 | Q98ia8 rhizobium l  | 222 | 7 | 0.9 | 216 | 17 | Q8TK92 | Q8tk92 methanosarc |
| 150 | 7   | 149  | 16 | Q92N96 | Q92n96 rhizobium m  | 223 | 7 | 0.9 | 217 | 16 | Q9CI71 | Q9ci71 lactococcus |
| 151 | 7   | 151  | 2  | Q9ANQ3 | Q9ang3 bradyrhizob  | 224 | 7 | 0.9 | 221 | 11 | Q8C6M4 | Q8c6m4 mus musculu |
| 152 | 7   | 152  | 13 | Q9DD88 | Q9dd88 oncorhynch   | 225 | 7 | 0.9 | 222 | 16 | Q984U9 | Q984u9 rhizobium l |
| 153 | 7   | 153  | 5  | Q00879 | Q00879 plasmodium   | 226 | 7 | 0.9 | 223 | 16 | Q9PMQ6 | Q9pmq6 campylobact |
| 154 | 7   | 153  | 10 | Q94F22 | Q94f22 arabidopsis  | 227 | 7 | 0.9 | 223 | 16 | Q9P4B6 | Q9p4b6 xanthomonas |
| 155 | 7   | 153  | 10 | Q9LY07 | Q9ly07 arabidopsis  | 228 | 7 | 0.9 | 226 | 16 | Q8XCM7 | Q8xcm7 escherichia |
| 156 | 7   | 156  | 16 | Q8YT08 | Q8ytg8 anabaena sp  | 229 | 7 | 0.9 | 228 | 5  | Q17275 | Q17275 brugia paha |
| 157 | 7   | 156  | 16 | Q8XSU0 | Q8xsu0 ralstonia s  | 230 | 7 | 0.9 | 228 | 16 | Q9A4A2 | Q9a4a2 caulobacter |
| 158 | 7   | 156  | 16 | Q8DKK2 | Q8dkr2 synecococc   | 231 | 7 | 0.9 | 231 | 16 | Q8XWU8 | Q8xwu8 ralstonia s |
| 159 | 7   | 158  | 2  | Q9RN96 | Q9rn96 streptococc  | 232 | 7 | 0.9 | 233 | 2  | Q8GGE5 | Q8gge5 streptomyc  |
| 160 | 7   | 158  | 16 | Q8PH21 | Q8ph21 xanthomonas  | 233 | 7 | 0.9 | 233 | 11 | Q925N7 | Q925n7 rattus norv |
| 161 | 7   | 159  | 16 | Q8EI22 | Q8ei22 shewanella   | 234 | 7 | 0.9 | 235 | 10 | Q8LMQ7 | Q8lmq7 oryza sativ |
| 162 | 7   | 161  | 16 | Q92KD9 | Q92kd9 rhizobium m  | 235 | 7 | 0.9 | 236 | 2  | Q9XHT4 | Q9xht4 pseudomonas |

|     |   |     |     |    |        |                     |     |   |     |     |    |         |
|-----|---|-----|-----|----|--------|---------------------|-----|---|-----|-----|----|---------|
| 236 | 7 | 0.9 | 236 | 12 | Q85028 | Q85028 pseudorabies | 309 | 7 | 0.9 | 300 | 8  | Q9MFC7  |
| 237 | 7 | 0.9 | 237 | 15 | Q8DBE9 | Q8dbb9 vibrio vuln  | 310 | 7 | 0.9 | 303 | 16 | Q98G68  |
| 238 | 7 | 0.9 | 237 | 5  | Q17242 | Q17242 vibrio mala  | 311 | 7 | 0.9 | 306 | 8  | Q79467  |
| 239 | 7 | 0.9 | 238 | 4  | Q8TAN7 | Q8tan7 homo sapien  | 312 | 7 | 0.9 | 306 | 11 | Q8CC62  |
| 240 | 7 | 0.9 | 238 | 11 | Q9COV7 | Q9cgv7 mus musculus | 313 | 7 | 0.9 | 306 | 16 | Q9KKT0  |
| 241 | 7 | 0.9 | 238 | 16 | Q9RW58 | Q9rw58 deirococcus  | 314 | 7 | 0.9 | 307 | 16 | Q8YDS6  |
| 242 | 7 | 0.9 | 242 | 2  | Q9EXK6 | Q9exk6 streptomyc   | 315 | 7 | 0.9 | 309 | 16 | Q9A2Z9  |
| 243 | 7 | 0.9 | 242 | 4  | Q8NHG8 | Q8nhg8 homo sapien  | 316 | 7 | 0.9 | 310 | 16 | Q9A2S4  |
| 244 | 7 | 0.9 | 242 | 9  | Q9MNC2 | Q9mcn2 bacterioph   | 317 | 7 | 0.9 | 312 | 7  | Q9X518  |
| 245 | 7 | 0.9 | 243 | 10 | Q9C7M8 | Q9c7m8 arabadopsis  | 318 | 7 | 0.9 | 314 | 11 | Q8VEV3  |
| 246 | 7 | 0.9 | 245 | 10 | Q9FUUV | Q9fuuv gonionomas   | 319 | 7 | 0.9 | 315 | 6  | Q95JP5  |
| 247 | 7 | 0.9 | 245 | 10 | Q9FUUV | Q9fuuv gonionomas   | 320 | 7 | 0.9 | 316 | 16 | Q92N15  |
| 248 | 7 | 0.9 | 245 | 10 | Q9SNW9 | Q9snw3 arabadopsis  | 321 | 7 | 0.9 | 318 | 2  | Q9KHQ2  |
| 249 | 7 | 0.9 | 245 | 10 | Q9FUUV | Q9fuuv gonionomas   | 322 | 7 | 0.9 | 318 | 10 | Q8H8P0  |
| 250 | 7 | 0.9 | 245 | 16 | Q45607 | Q45607 bacillus su  | 323 | 7 | 0.9 | 320 | 16 | Q9U0Q6  |
| 251 | 7 | 0.9 | 245 | 16 | Q8CQL4 | Q8cql4 staphylococ  | 324 | 7 | 0.9 | 320 | 2  | Q8RMK7  |
| 252 | 7 | 0.9 | 248 | 5  | Q9VBN2 | Q9vbn2 drosophila   | 325 | 7 | 0.9 | 323 | 2  | Q54490  |
| 253 | 7 | 0.9 | 249 | 16 | Q98GG5 | Q98gg5 thizobium l  | 326 | 7 | 0.9 | 323 | 5  | Q9V504  |
| 254 | 7 | 0.9 | 250 | 16 | Q9CHL0 | Q9chlo lactococcus  | 327 | 7 | 0.9 | 328 | 16 | Q8PFT5  |
| 255 | 7 | 0.9 | 255 | 10 | Q9LGN5 | Q9lgn5 oryza sativ  | 328 | 7 | 0.9 | 329 | 16 | Q9CEX8  |
| 256 | 7 | 0.9 | 256 | 2  | Q45893 | Q4589 alcaligenes   | 329 | 7 | 0.9 | 329 | 16 | Q9CEX8  |
| 257 | 7 | 0.9 | 256 | 12 | Q9JG85 | Q9jg85 tt virus.. o | 330 | 7 | 0.9 | 329 | 16 | Q8P7Q5  |
| 258 | 7 | 0.9 | 256 | 12 | Q9JG83 | Q9jg83 tt virus.. o | 331 | 7 | 0.9 | 330 | 16 | Q8DSC4  |
| 259 | 7 | 0.9 | 257 | 10 | Q9AU70 | Q9auz0 arabadopsis  | 332 | 7 | 0.9 | 333 | 16 | Q8PUJ3  |
| 260 | 7 | 0.9 | 257 | 11 | Q8BZ04 | Q8bz04 mus musculus | 333 | 7 | 0.9 | 333 | 16 | Q8P5L5  |
| 261 | 7 | 0.9 | 258 | 16 | Q8YJT7 | Q8yjt7 anabaena sp  | 334 | 7 | 0.9 | 335 | 2  | Q31146  |
| 262 | 7 | 0.9 | 258 | 16 | Q8P8G5 | Q8p8g5 xanthomonas  | 335 | 7 | 0.9 | 335 | 10 | Q8GXC9  |
| 263 | 7 | 0.9 | 261 | 3  | Q9P8P1 | Q9p8p1 glomus moss  | 336 | 7 | 0.9 | 335 | 16 | Q8SBC6  |
| 264 | 7 | 0.9 | 261 | 3  | Q9P8P9 | Q9p8p9 glomus moss  | 337 | 7 | 0.9 | 337 | 5  | Q9VG70  |
| 265 | 7 | 0.9 | 262 | 3  | Q9P8E8 | Q9p8e8 glomus moss  | 338 | 7 | 0.9 | 337 | 16 | Q8TQ82  |
| 266 | 7 | 0.9 | 265 | 16 | Q9AD22 | Q9ad22 streptomyc   | 339 | 7 | 0.9 | 339 | 17 | Q8TND5  |
| 267 | 7 | 0.9 | 268 | 16 | Q9A2W2 | Q9a2w2 caulobacter  | 340 | 7 | 0.9 | 340 | 10 | Q941H8  |
| 268 | 7 | 0.9 | 270 | 4  | Q8NGM6 | Q8ngm6 homo sapien  | 341 | 7 | 0.9 | 341 | 5  | Q81PD5  |
| 269 | 7 | 0.9 | 270 | 16 | Q9A3H3 | Q9a3h3 caulobacter  | 342 | 7 | 0.9 | 341 | 10 | Q49866  |
| 270 | 7 | 0.9 | 270 | 16 | Q9NLB1 | Q9nlb1 corynebacter | 343 | 7 | 0.9 | 343 | 16 | Q8XLJ9  |
| 271 | 7 | 0.9 | 271 | 11 | Q9VHC5 | Q9vhc5 mus musculu  | 344 | 7 | 0.9 | 343 | 16 | Q8P6J8  |
| 272 | 7 | 0.9 | 272 | 16 | Q9CD40 | Q9cd40 mycobacteri  | 345 | 7 | 0.9 | 343 | 16 | Q9XKZ3  |
| 273 | 7 | 0.9 | 272 | 16 | Q8PLV9 | Q8plv9 xanthomonas  | 346 | 7 | 0.9 | 344 | 3  | Q9P6W5  |
| 274 | 7 | 0.9 | 273 | 16 | P72047 | P72047 mycobacteri  | 347 | 7 | 0.9 | 345 | 4  | Q961H2  |
| 275 | 7 | 0.9 | 276 | 16 | Q92B87 | Q92b87 listeria in  | 348 | 7 | 0.9 | 345 | 16 | P73123  |
| 276 | 7 | 0.9 | 276 | 16 | Q8V6R2 | Q8v6r2 listeria mo  | 349 | 7 | 0.9 | 346 | 6  | Q95K56  |
| 277 | 7 | 0.9 | 276 | 16 | Q8V6R2 | Q8v6r2 listeria mo  | 349 | 7 | 0.9 | 346 | 8  | Q8W7B0  |
| 278 | 7 | 0.9 | 276 | 16 | Q8XS82 | Q8xs82 ralsconia s  | 350 | 7 | 0.9 | 346 | 8  | Q8W7B1  |
| 279 | 7 | 0.9 | 277 | 16 | Q8EF09 | Q8ef09 shewanella   | 351 | 7 | 0.9 | 347 | 3  | Q9HG11  |
| 280 | 7 | 0.9 | 277 | 10 | Q941H9 | Q941h9 nicotiana t  | 352 | 7 | 0.9 | 347 | 16 | Q8PG04  |
| 281 | 7 | 0.9 | 278 | 12 | Q83418 | Q83418 pseudorabie  | 353 | 7 | 0.9 | 347 | 17 | Q8U2F2  |
| 282 | 7 | 0.9 | 278 | 16 | Q9PHJ2 | Q9phj2 xylella fas  | 354 | 7 | 0.9 | 350 | 16 | Q9AAG8  |
| 283 | 7 | 0.9 | 279 | 16 | Q8E2L2 | Q8e2l2 streptococc  | 355 | 7 | 0.9 | 351 | 10 | Q8GUTH9 |
| 284 | 7 | 0.9 | 279 | 16 | Q8DMW3 | Q8dmw3 streptococ   | 356 | 7 | 0.9 | 352 | 2  | Q8B316  |
| 285 | 7 | 0.9 | 280 | 16 | Q8K6J2 | Q8k6j2 bacillus ha  | 357 | 7 | 0.9 | 352 | 16 | Q9A6L0  |
| 286 | 7 | 0.9 | 280 | 16 | Q8P9E9 | Q8p9e9 xanthomonas  | 358 | 7 | 0.9 | 352 | 16 | Q8U0H0  |
| 287 | 7 | 0.9 | 285 | 2  | Q52323 | Q52323 escherichia  | 359 | 7 | 0.9 | 353 | 16 | Q8U6D1  |
| 288 | 7 | 0.9 | 286 | 2  | Q9L8G4 | Q9lbg4 amycolatops  | 360 | 7 | 0.9 | 354 | 16 | Q97NG0  |
| 289 | 7 | 0.9 | 286 | 16 | Q8EF03 | Q8ef03 shewanella   | 361 | 7 | 0.9 | 354 | 16 | Q8Y094  |
| 290 | 7 | 0.9 | 291 | 10 | Q9M134 | Q9m134 arabadopsis  | 362 | 7 | 0.9 | 355 | 10 | Q9LE25  |
| 291 | 7 | 0.9 | 291 | 16 | Q8PM85 | Q8pm85 xanthomonas  | 363 | 7 | 0.9 | 360 | 5  | Q9V711  |
| 292 | 7 | 0.9 | 291 | 16 | Q8PA14 | Q8pa14 xanthomonas  | 364 | 7 | 0.9 | 360 | 8  | Q98864  |
| 293 | 7 | 0.9 | 292 | 4  | Q9NUN2 | Q9nun2 homo sapien  | 365 | 7 | 0.9 | 360 | 8  | Q8SM19  |
| 294 | 7 | 0.9 | 292 | 11 | Q9VECO | Q9veco mus musculu  | 366 | 7 | 0.9 | 361 | 8  | Q02686  |
| 295 | 7 | 0.9 | 293 | 10 | Q9RSP0 | Q9rsp0 oryza sativ  | 367 | 7 | 0.9 | 363 | 10 | Q8RVZ8  |
| 296 | 7 | 0.9 | 294 | 2  | Q9S5B6 | Q9s5b6 desulfovibr  | 368 | 7 | 0.9 | 363 | 10 | Q9SLY4  |
| 297 | 7 | 0.9 | 294 | 4  | Q9UF11 | Q9uf11 homo sapien  | 369 | 7 | 0.9 | 363 | 10 | Q9SLY5  |
| 298 | 7 | 0.9 | 294 | 16 | Q8XUJ9 | Q8xuj9 ralsconia s  | 370 | 7 | 0.9 | 365 | 11 | Q921U2  |
| 299 | 7 | 0.9 | 295 | 2  | Q8RPJ6 | Q8rpj6 desulfitoba  | 371 | 7 | 0.9 | 368 | 5  | Q9XYS5  |
| 300 | 7 | 0.9 | 295 | 10 | Q94B66 | Q94b66 arabadopsis  | 372 | 7 | 0.9 | 369 | 2  | Q9LB11  |
| 301 | 7 | 0.9 | 295 | 12 | Q9WB13 | Q9wb13 chilo iride  | 373 | 7 | 0.9 | 372 | 16 | Q9JVF9  |
| 302 | 7 | 0.9 | 295 | 16 | Q9X234 | Q9x234 thermotoga   | 374 | 7 | 0.9 | 374 | 10 | Q9LZP0  |
| 303 | 7 | 0.9 | 297 | 10 | Q9LU42 | Q9lu42 arabadopsis  | 375 | 7 | 0.9 | 375 | 10 | Q81435  |
| 304 | 7 | 0.9 | 297 | 10 | Q8LF08 | Q8lf08 arabadopsis  | 376 | 7 | 0.9 | 375 | 10 | P93066  |
| 305 | 7 | 0.9 | 297 | 16 | Q9C2P8 | Q9cp28 pasteuralla  | 377 | 7 | 0.9 | 375 | 5  | Q9N8E5  |
| 306 | 7 | 0.9 | 297 | 16 | Q9KYF2 | Q9kyf2 streptomyc   | 378 | 7 | 0.9 | 377 | 2  | Q9L3K8  |
| 307 | 7 | 0.9 | 299 | 2  | Q9L3A3 | Q9l3a3 sphingomona  | 379 | 7 | 0.9 | 377 | 2  | Q9L3K6  |
| 308 | 7 | 0.9 | 299 | 16 | Q8D4T3 | Q8d4t3 vibrio vuln  | 380 | 7 | 0.9 | 378 | 10 | Q94C35  |
| 309 | 7 | 0.9 | 300 | 2  | Q07868 | Q07868 streptococc  | 381 | 7 | 0.9 | 378 | 10 | Q9FF16  |



|     |   |     |     |    |        |                      |     |   |     |     |    |        |                     |
|-----|---|-----|-----|----|--------|----------------------|-----|---|-----|-----|----|--------|---------------------|
| 382 | 7 | 0.9 | 379 | 2  | Q9G376 | Q9G376 helicobacte   | 455 | 7 | 0.9 | 417 | 8  | Q9G2V4 | Q9G2v4 orthodicran  |
| 383 | 7 | 0.9 | 379 | 16 | Q25568 | Q25568 helicobacte   | 456 | 7 | 0.9 | 417 | 8  | Q9G2R3 | Q9G2r3 shariella    |
| 384 | 7 | 0.9 | 380 | 4  | Q95035 | Q95035 homo sapien   | 457 | 7 | 0.9 | 417 | 8  | Q9G2N7 | Q9G2n7 ulsta crisp  |
| 385 | 7 | 0.9 | 380 | 10 | Q04596 | Q04596 arabidopsis   | 458 | 7 | 0.9 | 417 | 8  | Q9G2V5 | Q9G2v5 orthodontiu  |
| 386 | 7 | 0.9 | 381 | 2  | Q91UR6 | Q91ur6 rhizobium m   | 459 | 7 | 0.9 | 417 | 8  | Q9G3M4 | Q9G3m4 arthodaea ri |
| 387 | 7 | 0.9 | 381 | 16 | Q92K79 | Q92kt9 helicobacte   | 460 | 7 | 0.9 | 417 | 11 | Q8B210 | Q8B210 mus musculu  |
| 388 | 7 | 0.9 | 385 | 16 | Q986C3 | Q986c3 rhizobium l   | 461 | 7 | 0.9 | 417 | 16 | Q9FC63 | Q9fc63 streptomyce  |
| 389 | 7 | 0.9 | 387 | 5  | Q18001 | Q18001 caenorhabdi   | 462 | 7 | 0.9 | 418 | 8  | Q94Z29 | Q94z29 peilia epip  |
| 390 | 7 | 0.9 | 388 | 4  | Q9NY42 | Q9ny42 homo sapien   | 463 | 7 | 0.9 | 419 | 16 | Q8EAF7 | Q8eaf7 shewanella   |
| 391 | 7 | 0.9 | 389 | 16 | Q98AK0 | Q98ak0 rhizobium l   | 464 | 7 | 0.9 | 420 | 16 | Q8RUE8 | Q8rue8 deinococcus  |
| 392 | 7 | 0.9 | 392 | 16 | Q9ZCJ5 | Q9zcj5 rickettsia    | 465 | 7 | 0.9 | 421 | 2  | Q9J177 | Q9j177 thermus the  |
| 393 | 7 | 0.9 | 393 | 5  | Q9XVP2 | Q9xvp2 caenorhabdi   | 466 | 7 | 0.9 | 421 | 8  | Q9G3L7 | Q9g3l7 atrichum un  |
| 394 | 7 | 0.9 | 393 | 16 | Q9K7L1 | Q9k7l1 bacillus ha   | 467 | 7 | 0.9 | 421 | 8  | Q9G2S2 | Q9g2s2 pogonatum u  |
| 395 | 7 | 0.9 | 394 | 5  | Q44002 | Q44002 toxoplasma    | 468 | 7 | 0.9 | 421 | 8  | Q9G2P2 | Q9g2p2 tetraphis p  |
| 396 | 7 | 0.9 | 394 | 16 | Q8NSY0 | Q8nsy0 corynebacte   | 469 | 7 | 0.9 | 423 | 8  | Q94Z26 | Q94z26 phaeoceros   |
| 397 | 7 | 0.9 | 395 | 10 | Q8RUK9 | Q8ruk9 oryza sativ   | 470 | 7 | 0.9 | 431 | 16 | Q8J3C6 | Q8j3c6 xanthomonas  |
| 398 | 7 | 0.9 | 399 | 16 | Q33Q25 | Q33q25 rhizobium m   | 471 | 7 | 0.9 | 432 | 16 | Q25371 | Q25371 helicobacte  |
| 399 | 7 | 0.9 | 401 | 16 | Q8XZAB | Q8xza8 ralestonia s  | 472 | 7 | 0.9 | 434 | 16 | Q9LH92 | Q9lh92 arabidopsis  |
| 400 | 7 | 0.9 | 403 | 16 | Q8NM62 | Q8nm62 corynebacte   | 473 | 7 | 0.9 | 434 | 16 | Q8F767 | Q8f767 leptospira   |
| 401 | 7 | 0.9 | 405 | 5  | Q17651 | Q17651 caenorhabdi   | 474 | 7 | 0.9 | 435 | 16 | Q8DB48 | Q8db48 vibrio vuln  |
| 402 | 7 | 0.9 | 405 | 16 | Q8UEJ6 | Q8uej6 agrobacteri   | 475 | 7 | 0.9 | 439 | 2  | Q8KL48 | Q8kl48 rhizobium e  |
| 403 | 7 | 0.9 | 409 | 16 | Q913D2 | Q913d2 pseudomonas   | 476 | 7 | 0.9 | 441 | 4  | Q75685 | Q75685 homo sapien  |
| 404 | 7 | 0.9 | 409 | 16 | Q8XRL2 | Q8xrl2 ralestonia s  | 477 | 7 | 0.9 | 442 | 12 | Q8JKN6 | Q8jkn6 heliothis z  |
| 405 | 7 | 0.9 | 410 | 16 | Q8YD86 | Q8yd86 bruceella me  | 478 | 7 | 0.9 | 444 | 16 | Q8CLK3 | Q8clk3 pasteurella  |
| 406 | 7 | 0.9 | 410 | 16 | Q8FV45 | Q8fv45 bruceella su  | 479 | 7 | 0.9 | 445 | 16 | Q9FKY9 | Q9fky9 chlamydia m  |
| 407 | 7 | 0.9 | 411 | 5  | Q22718 | Q22718 caenorhabdi   | 480 | 7 | 0.9 | 445 | 16 | Q97DQ0 | Q97dq0 clostridium  |
| 408 | 7 | 0.9 | 411 | 5  | Q8I431 | Q8i431 plasmodium    | 481 | 7 | 0.9 | 446 | 4  | Q60650 | Q60650 homo sapien  |
| 409 | 7 | 0.9 | 412 | 2  | Q95595 | Q95595 rhodobacter   | 482 | 7 | 0.9 | 451 | 11 | Q9QUN4 | Q9qun4 rattus norv  |
| 410 | 7 | 0.9 | 412 | 4  | Q14184 | Q14184 homo sapien   | 483 | 7 | 0.9 | 452 | 17 | Q29870 | Q29870 archaeoglob  |
| 411 | 7 | 0.9 | 412 | 16 | Q9AE66 | Q9ae66 caulobacter   | 484 | 7 | 0.9 | 453 | 2  | Q8XP13 | Q8xp13 bartonella   |
| 412 | 7 | 0.9 | 414 | 8  | Q956P4 | Q956p4 anthoceros    | 485 | 7 | 0.9 | 457 | 16 | Q53893 | Q53893 mycobacteri  |
| 413 | 7 | 0.9 | 415 | 2  | Q2LAY7 | Q2lay7 streptococc   | 486 | 7 | 0.9 | 461 | 11 | Q9D168 | Q9d168 mus musculu  |
| 414 | 7 | 0.9 | 416 | 16 | Q99X43 | Q99x43 staphylococ   | 487 | 7 | 0.9 | 462 | 16 | Q8PIB3 | Q8piB3 xanthomonas  |
| 415 | 7 | 0.9 | 417 | 8  | Q94ZK2 | Q94zk2 corsinia co   | 488 | 7 | 0.9 | 465 | 11 | Q9QWF3 | Q9qwf3 spermophili  |
| 416 | 7 | 0.9 | 417 | 8  | Q9G3A8 | Q9g3a8 buxbaumia a   | 489 | 7 | 0.9 | 465 | 11 | Q88354 | Q88354 spermophili  |
| 417 | 7 | 0.9 | 417 | 8  | Q9G2W5 | Q9g2w5 mniun hornu   | 490 | 7 | 0.9 | 466 | 17 | Q9PJV3 | Q9pvj3 methanosarc  |
| 418 | 7 | 0.9 | 417 | 8  | Q94ZAB | Q94za8 lunularia c   | 491 | 7 | 0.9 | 469 | 16 | Q9XA62 | Q9xa62 streptomyce  |
| 419 | 7 | 0.9 | 417 | 8  | Q9G340 | Q9g340 leskea poly   | 492 | 7 | 0.9 | 472 | 2  | Q9LA86 | Q9la86 aeromonas h  |
| 420 | 7 | 0.9 | 417 | 8  | Q94Z32 | Q94z32 piagioclila   | 493 | 7 | 0.9 | 472 | 16 | Q8YQJ6 | Q8yqj6 anabaena sp  |
| 421 | 7 | 0.9 | 417 | 8  | Q9G2P4 | Q9g2p4 takaria lep   | 494 | 7 | 0.9 | 473 | 8  | Q9TKZ9 | Q9tkz9 nephroselm   |
| 422 | 7 | 0.9 | 417 | 8  | Q9G2R7 | Q9g2r7 sphagnum fa   | 495 | 7 | 0.9 | 474 | 16 | Q53750 | Q53750 mycobacteri  |
| 423 | 7 | 0.9 | 417 | 8  | Q9G2R4 | Q9g2r4 scorpidium    | 496 | 7 | 0.9 | 475 | 10 | Q9VZ21 | Q9vz21 arabidopsis  |
| 424 | 7 | 0.9 | 417 | 8  | Q9G2R1 | Q9g2r1 timmia bava   | 497 | 7 | 0.9 | 475 | 10 | Q8RXP7 | Q8rx7 arabidopsis   |
| 425 | 7 | 0.9 | 417 | 8  | Q9G2T9 | Q9g2t9 pterogonium   | 498 | 7 | 0.9 | 476 | 16 | Q8VKA5 | Q8vka5 mycobacteri  |
| 426 | 7 | 0.9 | 417 | 8  | Q9G2P5 | Q9g2p5 tortula lat   | 499 | 7 | 0.9 | 478 | 16 | Q9S2I9 | Q9s2i9 streptomyce  |
| 427 | 7 | 0.9 | 417 | 8  | Q9G343 | Q9g343 isothecium    | 500 | 7 | 0.9 | 479 | 5  | Q9U1V6 | Q9u1v6 caenorhabdi  |
| 428 | 7 | 0.9 | 417 | 8  | Q9G398 | Q9g398 dichodontiu   | 501 | 7 | 0.9 | 479 | 10 | Q8LST7 | Q8lst7 chlamydomon  |
| 429 | 7 | 0.9 | 417 | 8  | Q9G2R2 | Q9g2r2 thamnobryum   | 502 | 7 | 0.9 | 480 | 16 | Q92JZ0 | Q92jz0 rhizobium m  |
| 430 | 7 | 0.9 | 417 | 8  | Q9G377 | Q9g377 fissidens c   | 503 | 7 | 0.9 | 481 | 16 | Q8UGV5 | Q8ugv5 agrobacteri  |
| 431 | 7 | 0.9 | 417 | 8  | Q9G6U4 | Q9g6u4 aulacomnium   | 504 | 7 | 0.9 | 482 | 16 | Q9L254 | Q9l254 streptomyce  |
| 432 | 7 | 0.9 | 417 | 8  | Q9G2R6 | Q9g2r6 plagiopopus o | 505 | 7 | 0.9 | 484 | 16 | Q8EF16 | Q8ef16 shewanella   |
| 433 | 7 | 0.9 | 417 | 8  | Q9G346 | Q9g346 hygrophynum   | 506 | 7 | 0.9 | 486 | 16 | Q8EX62 | Q8ex62 leptospira   |
| 434 | 7 | 0.9 | 417 | 8  | Q94ZM1 | Q94zm1 bucegia rom   | 507 | 7 | 0.9 | 488 | 8  | Q21360 | Q21360 triticum ae  |
| 435 | 7 | 0.9 | 417 | 8  | Q9G2R9 | Q9g2r9 zinacocarpus  | 508 | 7 | 0.9 | 488 | 8  | Q9MF79 | Q9mf79 beta vulgar  |
| 436 | 7 | 0.9 | 417 | 8  | Q94ZC8 | Q94zc8 fessombroni   | 509 | 7 | 0.9 | 488 | 8  | Q21359 | Q21359 triticum ae  |
| 437 | 7 | 0.9 | 417 | 8  | Q9G3A7 | Q9g3a7 bartramia h   | 510 | 7 | 0.9 | 488 | 8  | Q8HCM1 | Q8hcm1 oryza sativ  |
| 438 | 7 | 0.9 | 417 | 8  | Q9G3A1 | Q9g3a1 ceratodon p   | 511 | 7 | 0.9 | 490 | 11 | Q9CTA9 | Q9cta9 mus musculu  |
| 439 | 7 | 0.9 | 417 | 8  | Q9G2S3 | Q9g2s3 pottia trun   | 512 | 7 | 0.9 | 491 | 16 | Q984W7 | Q984w7 rhizobium l  |
| 440 | 7 | 0.9 | 417 | 8  | Q9G348 | Q9g348 hedwigia ci   | 513 | 7 | 0.9 | 491 | 16 | Q984P4 | Q984p4 drosophila   |
| 441 | 7 | 0.9 | 417 | 8  | Q9G399 | Q9g399 ditrichum c   | 514 | 7 | 0.9 | 493 | 5  | Q8SZB1 | Q8szr1 drosophila   |
| 442 | 7 | 0.9 | 417 | 8  | Q9G2T1 | Q9g2t1 physcomitre   | 515 | 7 | 0.9 | 493 | 11 | Q8CE55 | Q8ce55 mus musculu  |
| 443 | 7 | 0.9 | 417 | 8  | Q9G397 | Q9g397 diphyscium    | 516 | 7 | 0.9 | 493 | 11 | Q8BQJ6 | Q8bqj6 mus musculu  |
| 444 | 7 | 0.9 | 417 | 8  | Q9G344 | Q9g344 homalia tri   | 517 | 7 | 0.9 | 494 | 12 | Q9DVT4 | Q9dvt4 plutella xy  |
| 445 | 7 | 0.9 | 417 | 8  | Q9G394 | Q9g394 encalypta s   | 518 | 7 | 0.9 | 496 | 10 | Q9S833 | Q9s833 arabidopsis  |
| 446 | 7 | 0.9 | 417 | 8  | Q9G341 | Q9g341 leucobryum    | 519 | 7 | 0.9 | 499 | 8  | Q95621 | Q95621 arabidopsis  |
| 447 | 7 | 0.9 | 417 | 8  | Q94YX0 | Q94yx0 sphaerocarpi  | 520 | 7 | 0.9 | 499 | 10 | Q05000 | Q05000 arabidopsis  |
| 448 | 7 | 0.9 | 417 | 8  | Q9G380 | Q9g380 funtinalis    | 521 | 7 | 0.9 | 505 | 10 | Q9SUE2 | Q9sue2 arabidopsis  |
| 449 | 7 | 0.9 | 417 | 8  | Q9G349 | Q9g349 funaria hyg   | 522 | 7 | 0.9 | 506 | 8  | Q47138 | Q47138 rhodotianu   |
| 450 | 7 | 0.9 | 417 | 8  | Q9G2R5 | Q9g2r5 schistocteg   | 523 | 7 | 0.9 | 506 | 8  | Q47150 | Q47150 phyllocoe    |
| 451 | 7 | 0.9 | 417 | 8  | Q9G2T2 | Q9g2t2 pohlia nuta   | 524 | 7 | 0.9 | 506 | 8  | Q47135 | Q47135 phyllocoe    |
| 452 | 7 | 0.9 | 417 | 8  | Q94YX7 | Q94yx7 ricciocarpus  | 525 | 7 | 0.9 | 506 | 8  | Q47142 | Q47142 phyllocoe    |
| 453 | 7 | 0.9 | 417 | 8  | Q9G3A0 | Q9g3a0 cinctidotus   | 526 | 7 | 0.9 | 508 | 5  | Q9NA84 | Q9na84 caenorhabdi  |
| 454 | 7 | 0.9 | 417 | 8  | Q9G2P3 | Q9g2p3 tomentypnum   | 527 | 7 | 0.9 | 508 | 8  | Q47140 | Q47140 kalmiopsis   |

|     |   |     |     |    |        |                    |     |   |     |     |    |         |                     |
|-----|---|-----|-----|----|--------|--------------------|-----|---|-----|-----|----|---------|---------------------|
| 528 | 7 | 0.9 | 509 | 1  | Q97M9  | Q97M9 uncultured   | 601 | 7 | 0.9 | 629 | 16 | Q8ZRT1  | Q8ZRT1 salmonella   |
| 529 | 7 | 0.9 | 510 | 16 | Q9PAY3 | Q9PAY3 xylella fas | 602 | 7 | 0.9 | 629 | 16 | Q8Z9B9  | Q8Z9B9 salmonella   |
| 530 | 7 | 0.9 | 511 | 5  | Q9VM14 | Q9VM14 drosophila  | 603 | 7 | 0.9 | 630 | 5  | Q23591  | Q23591 caenorhabdi  |
| 531 | 7 | 0.9 | 512 | 4  | Q9NWD4 | Q9NWD4 homo sapien | 604 | 7 | 0.9 | 630 | 16 | Q8X966  | Q8X966 escherichia  |
| 532 | 7 | 0.9 | 513 | 10 | Q9SOY6 | Q9SOY6 arabidopsis | 605 | 7 | 0.9 | 631 | 16 | Q8DKW7  | Q8DKW7 synechococ   |
| 533 | 7 | 0.9 | 514 | 10 | Q9FLS7 | Q9FLS7 arabidopsis | 606 | 7 | 0.9 | 632 | 10 | Q9SM30  | Q9SM30 taraxacum o  |
| 534 | 7 | 0.9 | 515 | 10 | Q94D22 | Q94D22 oryza sativ | 607 | 7 | 0.9 | 634 | 5  | Q17474  | Q17474 caenorhabdi  |
| 535 | 7 | 0.9 | 516 | 10 | Q9ASQ7 | Q9ASQ7 arabidopsis | 608 | 7 | 0.9 | 634 | 5  | Q17475  | Q17475 caenorhabdi  |
| 536 | 7 | 0.9 | 517 | 10 | Q9EM81 | Q9EM81 rhizobium l | 609 | 7 | 0.9 | 636 | 10 | Q8RVM4  | Q8RVM4 lycopersico  |
| 537 | 7 | 0.9 | 518 | 12 | Q8ZJ67 | Q8ZJ67 versinia pe | 610 | 7 | 0.9 | 637 | 10 | Q23786  | Q23786 cynara scol  |
| 538 | 7 | 0.9 | 519 | 12 | Q83417 | Q83417 pseudorabie | 611 | 7 | 0.9 | 637 | 4  | Q9H7T5  | Q9H7T5 homo sapien  |
| 539 | 7 | 0.9 | 520 | 16 | Q8DJA1 | Q8DJA1 synechococ  | 612 | 7 | 0.9 | 639 | 4  | Q9ABR1  | Q9ABR1 caulobacter  |
| 540 | 7 | 0.9 | 521 | 16 | Q9RYE8 | Q9RYE8 deinococcc  | 613 | 7 | 0.9 | 641 | 16 | Q9ABR1  | Q9ABR1 caulobacter  |
| 541 | 7 | 0.9 | 522 | 12 | Q8S527 | Q8S527 pseudorabie | 614 | 7 | 0.9 | 643 | 3  | Q42633  | Q42633 cochllobola  |
| 542 | 7 | 0.9 | 523 | 12 | Q8GBD0 | Q8GBD0 enterococcc | 615 | 7 | 0.9 | 649 | 6  | Q8KLE4  | Q8KLE4 mycoplasma   |
| 543 | 7 | 0.9 | 524 | 2  | Q8GBD6 | Q8GBD6 enterococcc | 616 | 7 | 0.9 | 649 | 6  | Q28657  | Q28657 cryptotagus  |
| 544 | 7 | 0.9 | 525 | 2  | Q8GBD4 | Q8GBD4 enterococcc | 617 | 7 | 0.9 | 656 | 16 | Q8YVM28 | Q8YVM28 anabaena sp |
| 545 | 7 | 0.9 | 526 | 2  | Q8GBD2 | Q8GBD2 enterococcc | 618 | 7 | 0.9 | 658 | 16 | Q8PR52  | Q8PR52 xanthomonas  |
| 546 | 7 | 0.9 | 527 | 2  | Q8GBD0 | Q8GBD0 enterococcc | 619 | 7 | 0.9 | 660 | 11 | Q8C7B6  | Q8C7B6 mus musculu  |
| 547 | 7 | 0.9 | 528 | 5  | Q9W2P0 | Q9W2P0 drosophila  | 620 | 7 | 0.9 | 662 | 16 | Q93RX7  | Q93RX7 streptomyce  |
| 548 | 7 | 0.9 | 529 | 16 | Q8EPE0 | Q8EPE0 shewanella  | 621 | 7 | 0.9 | 663 | 5  | Q9VQS3  | Q9VQS3 drosophila   |
| 549 | 7 | 0.9 | 530 | 16 | Q9RK71 | Q9RK71 deinococcc  | 622 | 7 | 0.9 | 663 | 5  | Q8S297  | Q8S297 drosophila   |
| 550 | 7 | 0.9 | 531 | 16 | Q8RC12 | Q8RC12 thermoanaer | 623 | 7 | 0.9 | 671 | 2  | Q9WVH5  | Q9WVH5 pseudomonas  |
| 551 | 7 | 0.9 | 532 | 16 | Q8G491 | Q8G491 bifidobacte | 624 | 7 | 0.9 | 671 | 2  | Q9WVH9  | Q9WVH9 pseudomonas  |
| 552 | 7 | 0.9 | 533 | 16 | Q8G491 | Q8G491 bifidobacte | 625 | 7 | 0.9 | 676 | 2  | Q46625  | Q46625 erwinia ste  |
| 553 | 7 | 0.9 | 534 | 16 | Q8JRX9 | Q8JRX9 chlamydia p | 626 | 7 | 0.9 | 677 | 2  | Q9FC21  | Q9FC21 erwinia ste  |
| 554 | 7 | 0.9 | 535 | 10 | Q8LMB8 | Q8LMB8 oryza sativ | 627 | 7 | 0.9 | 677 | 8  | Q99982  | Q99982 porphyra pu  |
| 555 | 7 | 0.9 | 536 | 10 | Q8RV21 | Q8RV21 deinococcc  | 628 | 7 | 0.9 | 681 | 4  | Q93X91  | Q93X91 homo sapien  |
| 556 | 7 | 0.9 | 537 | 4  | Q9NSA9 | Q9NSA9 homo sapien | 629 | 7 | 0.9 | 681 | 4  | Q8TAL3  | Q8TAL3 homo sapien  |
| 557 | 7 | 0.9 | 538 | 4  | Q9SE78 | Q9SE78 homo sapien | 630 | 7 | 0.9 | 681 | 5  | Q9XUS9  | Q9XUS9 caenorhabdi  |
| 558 | 7 | 0.9 | 539 | 16 | Q9KUS2 | Q9KUS2 streptomyce | 631 | 7 | 0.9 | 686 | 5  | Q9V550  | Q9V550 drosophila   |
| 559 | 7 | 0.9 | 540 | 16 | Q8PJH9 | Q8PJH9 xanthomonas | 632 | 7 | 0.9 | 686 | 16 | Q8F208  | Q8F208 leptospira   |
| 560 | 7 | 0.9 | 541 | 16 | Q8KY42 | Q8KY42 streptomyce | 633 | 7 | 0.9 | 687 | 10 | Q9LXF7  | Q9LXF7 arabidopsis  |
| 561 | 7 | 0.9 | 542 | 16 | Q961R9 | Q961R9 drosophila  | 634 | 7 | 0.9 | 689 | 2  | Q8RT13  | Q8RT13 pseudalter   |
| 562 | 7 | 0.9 | 543 | 16 | Q92924 | Q92924 chlamydia p | 635 | 7 | 0.9 | 690 | 5  | Q9XUC6  | Q9XUC6 caenorhabdi  |
| 563 | 7 | 0.9 | 544 | 2  | Q45632 | Q45632 caenorhabdi | 636 | 7 | 0.9 | 690 | 10 | Q8H6B4  | Q8H6B4 marchantia   |
| 564 | 7 | 0.9 | 545 | 2  | Q8VW86 | Q8VW86 streptomyce | 637 | 7 | 0.9 | 690 | 16 | Q9PSB6  | Q9PSB6 xanthomonas  |
| 565 | 7 | 0.9 | 546 | 12 | Q8UBT8 | Q8UBT8 homo sapien | 638 | 7 | 0.9 | 692 | 10 | Q8L644  | Q8L644 arabidopsis  |
| 566 | 7 | 0.9 | 547 | 2  | Q68522 | Q68522 myxococcus  | 639 | 7 | 0.9 | 692 | 16 | Q8PPJ1  | Q8PPJ1 xanthomonas  |
| 567 | 7 | 0.9 | 548 | 16 | Q8PUX7 | Q8PUX7 agrobacteri | 640 | 7 | 0.9 | 693 | 10 | Q99018  | Q99018 capsicum an  |
| 568 | 7 | 0.9 | 549 | 16 | Q8P300 | Q8P300 xanthomonas | 641 | 7 | 0.9 | 698 | 5  | Q21341  | Q21341 caenorhabdi  |
| 569 | 7 | 0.9 | 550 | 10 | Q9Z332 | Q9Z332 daucus caro | 642 | 7 | 0.9 | 699 | 16 | Q9ABL0  | Q9ABL0 caulobacter  |
| 570 | 7 | 0.9 | 551 | 10 | Q9L072 | Q9L072 arabidopsis | 643 | 7 | 0.9 | 700 | 2  | Q87441  | Q87441 pseudomonas  |
| 571 | 7 | 0.9 | 552 | 16 | Q8G479 | Q8G479 bifidobacte | 644 | 7 | 0.9 | 702 | 12 | Q89681  | Q89681 vesicular e  |
| 572 | 7 | 0.9 | 553 | 10 | Q93YW3 | Q93YW3 arabidopsis | 645 | 7 | 0.9 | 702 | 11 | Q9D6Y9  | Q9D6Y9 mus musculu  |
| 573 | 7 | 0.9 | 554 | 16 | Q9KZ09 | Q9KZ09 streptomyce | 646 | 7 | 0.9 | 706 | 4  | Q96SX2  | Q96SX2 homo sapien  |
| 574 | 7 | 0.9 | 555 | 16 | Q9KXV0 | Q9KXV0 vibrio chol | 647 | 7 | 0.9 | 706 | 4  | Q9BX90  | Q9BX90 homo sapien  |
| 575 | 7 | 0.9 | 556 | 5  | Q8IKD1 | Q8IKD1 plasmodium  | 648 | 7 | 0.9 | 711 | 10 | Q8Z304  | Q8Z304 arabidopsis  |
| 576 | 7 | 0.9 | 557 | 5  | Q20517 | Q20517 caenorhabdi | 649 | 7 | 0.9 | 711 | 10 | Q8RX22  | Q8RX22 arabidopsis  |
| 577 | 7 | 0.9 | 558 | 12 | Q8QPT4 | Q8QPT4 chimpanzee  | 650 | 7 | 0.9 | 712 | 11 | Q99K34  | Q99K34 mus musculu  |
| 578 | 7 | 0.9 | 559 | 5  | Q9U750 | Q9U750 plasmodium  | 651 | 7 | 0.9 | 713 | 2  | Q93PY6  | Q93PY6 pseudomonas  |
| 579 | 7 | 0.9 | 560 | 5  | Q8IDR1 | Q8IDR1 plasmodium  | 652 | 7 | 0.9 | 713 | 16 | Q8PJ00  | Q8PJ00 xanthomonas  |
| 580 | 7 | 0.9 | 561 | 10 | Q94017 | Q94017 arabidopsis | 653 | 7 | 0.9 | 716 | 2  | Q8KQY3  | Q8KQY3 vibrio chol  |
| 581 | 7 | 0.9 | 562 | 13 | Q919B3 | Q919B3 xenopus lae | 654 | 7 | 0.9 | 716 | 2  | Q8RL15  | Q8RL15 providencia  |
| 582 | 7 | 0.9 | 563 | 5  | Q17935 | Q17935 caenorhabdi | 655 | 7 | 0.9 | 719 | 5  | Q27365  | Q27365 caenorhabdi  |
| 583 | 7 | 0.9 | 564 | 16 | Q8DCM2 | Q8DCM2 vibrio vuln | 656 | 7 | 0.9 | 722 | 5  | Q21448  | Q21448 caenorhabdi  |
| 584 | 7 | 0.9 | 565 | 4  | Q9H6W1 | Q9H6W1 homo sapien | 657 | 7 | 0.9 | 723 | 16 | Q9KXN1  | Q9KXN1 vibrio chol  |
| 585 | 7 | 0.9 | 566 | 16 | Q9S2Q5 | Q9S2Q5 streptomyce | 658 | 7 | 0.9 | 725 | 17 | Q97Z97  | Q97Z97 sulfolobus   |
| 586 | 7 | 0.9 | 567 | 16 | Q8F802 | Q8F802 leptospira  | 659 | 7 | 0.9 | 729 | 5  | Q9GS25  | Q9GS25 drosophila   |
| 587 | 7 | 0.9 | 568 | 16 | Q8DGN3 | Q8DGN3 synechococ  | 660 | 7 | 0.9 | 735 | 5  | Q965H3  | Q965H3 caenorhabdi  |
| 588 | 7 | 0.9 | 569 | 5  | Q810J6 | Q810J6 dictyosteli | 661 | 7 | 0.9 | 738 | 10 | Q8L7B1  | Q8L7B1 arabidopsis  |
| 589 | 7 | 0.9 | 570 | 16 | Q8XHL3 | Q8XHL3 clostridium | 662 | 7 | 0.9 | 739 | 10 | Q9LQHO  | Q9LQHO arabidopsis  |
| 590 | 7 | 0.9 | 571 | 5  | Q20798 | Q20798 caenorhabdi | 663 | 7 | 0.9 | 749 | 11 | Q9QXK0  | Q9QXK0 rattus norv  |
| 591 | 7 | 0.9 | 572 | 12 | Q71123 | Q71123 rhesus cyto | 664 | 7 | 0.9 | 749 | 11 | Q9D323  | Q9D323 mus musculu  |
| 592 | 7 | 0.9 | 573 | 5  | Q17936 | Q17936 caenorhabdi | 665 | 7 | 0.9 | 749 | 11 | Q8C497  | Q8C497 mus musculu  |
| 593 | 7 | 0.9 | 574 | 12 | Q9Z614 | Q9Z614 streptomyce | 666 | 7 | 0.9 | 749 | 11 | Q8C3V4  | Q8C3V4 mus musculu  |
| 594 | 7 | 0.9 | 575 | 5  | Q27334 | Q27334 onchocerca  | 667 | 7 | 0.9 | 755 | 11 | Q8C8M3  | Q8C8M3 mus musculu  |
| 595 | 7 | 0.9 | 576 | 2  | Q9EXH3 | Q9EXH3 listeria iv | 668 | 7 | 0.9 | 757 | 2  | Q8VRL4  | Q8VRL4 thermus the  |
| 596 | 7 | 0.9 | 577 | 10 | Q8LJ11 | Q8LJ11 oryza sativ | 669 | 7 | 0.9 | 757 | 10 | Q93WU1  | Q93WU1 avena strig  |
| 597 | 7 | 0.9 | 578 | 16 | Q98LH3 | Q98LH3 rhizobium l | 670 | 7 | 0.9 | 763 | 5  | Q9V114  | Q9V114 drosophila   |
| 598 | 7 | 0.9 | 579 | 4  | Q9H943 | Q9H943 homo sapien | 671 | 7 | 0.9 | 765 | 12 | Q9QUE1  | Q9QUE1 tt virus. h  |
| 599 | 7 | 0.9 | 580 | 16 | Q8VPL6 | Q8VPL6 anabaena sp | 672 | 7 | 0.9 | 765 | 12 | Q9JGS5  | Q9JGS5 tt virus. p  |
| 600 | 7 | 0.9 | 581 | 16 | Q8D3K0 | Q8D3K0 wigleswort  | 673 | 7 | 0.9 | 765 | 12 | Q9JGS8  | Q9JGS8 tt virus. p  |

|     |   |     |      |    |        |                     |     |   |     |      |    |        |                    |
|-----|---|-----|------|----|--------|---------------------|-----|---|-----|------|----|--------|--------------------|
| 674 | 7 | 0.9 | 765  | 12 | Q9JGS2 | Q9jgs2 tt virus. p  | 747 | 7 | 0.9 | 1106 | 16 | Q8F4S3 | Q8f4s3 leptospira  |
| 675 | 7 | 0.9 | 767  | 5  | Q03625 | Q03625 trypanosoma  | 748 | 7 | 0.9 | 1117 | 10 | Q9LRY7 | Q9lry7 arabidopsis |
| 676 | 7 | 0.9 | 771  | 10 | Q23326 | Q23326 arabidopsis  | 749 | 7 | 0.9 | 1130 | 5  | Q8MNP8 | Q8mnp8 dictyosteli |
| 677 | 7 | 0.9 | 775  | 10 | Q9LG72 | Q9lg72 oryza sativ  | 750 | 7 | 0.9 | 1132 | 5  | Q8MRQ9 | Q8mrq9 drosophila  |
| 678 | 7 | 0.9 | 776  | 11 | Q8KJN1 | Q8kjin mus musculus | 751 | 7 | 0.9 | 1132 | 16 | Q9RRQ7 | Q9rrq7 deinococcus |
| 679 | 7 | 0.9 | 781  | 16 | Q8PJU9 | Q8pjju xanthomonas  | 752 | 7 | 0.9 | 1134 | 16 | Q92RW9 | Q92rw9 rhizobium m |
| 680 | 7 | 0.9 | 782  | 4  | Q9NP80 | Q9np80 homo sapien  | 753 | 7 | 0.9 | 1135 | 5  | Q9NUQ4 | Q9nuq4 paramedium  |
| 681 | 7 | 0.9 | 786  | 5  | Q8T1B7 | Q8tlb7 dictyosteli  | 754 | 7 | 0.9 | 1147 | 10 | Q9SA26 | Q9sa26 arabidopsis |
| 682 | 7 | 0.9 | 789  | 16 | Q8XX15 | Q8xx15 ralstonia s  | 755 | 7 | 0.9 | 1165 | 5  | Q9BJ47 | Q9bj47 leishmania  |
| 683 | 7 | 0.9 | 796  | 16 | Q924W5 | Q924w5 streptomyce  | 756 | 7 | 0.9 | 1171 | 3  | Q9P3E2 | Q9pxu1 drosophila  |
| 684 | 7 | 0.9 | 802  | 5  | Q969B9 | Q969b9 giardia lam  | 757 | 7 | 0.9 | 1208 | 5  | Q9VKU1 | Q9vk44 drosophila  |
| 685 | 7 | 0.9 | 802  | 17 | Q97XQ7 | Q97xq7 sulfolobus   | 758 | 7 | 0.9 | 1217 | 5  | Q9VK44 | Q9vk44 xenopus lae |
| 686 | 7 | 0.9 | 803  | 11 | Q9DC20 | Q9dc20 mus musculus | 759 | 7 | 0.9 | 1219 | 13 | Q73749 | Q73749 xenopus lae |
| 687 | 7 | 0.9 | 808  | 10 | Q8H7E6 | Q8h7e6 arabidopsis  | 760 | 7 | 0.9 | 1248 | 10 | Q9SAJ2 | Q9saj2 arabidopsis |
| 688 | 7 | 0.9 | 809  | 10 | Q94G49 | Q94g49 arabidopsis  | 761 | 7 | 0.9 | 1248 | 16 | Q8UIO1 | Q8ui01 agrobacteri |
| 689 | 7 | 0.9 | 813  | 10 | Q9LQ00 | Q9lqu0 arabidopsis  | 762 | 7 | 0.9 | 1256 | 10 | Q23528 | Q23528 arabidopsis |
| 690 | 7 | 0.9 | 814  | 5  | Q965Y0 | Q965y0 caenorhabdi  | 763 | 7 | 0.9 | 1272 | 5  | Q8MRK2 | Q8mrk2 drosophila  |
| 691 | 7 | 0.9 | 815  | 16 | Q8ZIM9 | Q8zim9 yerseinia pe | 764 | 7 | 0.9 | 1272 | 16 | Q9FBR4 | Q9fbr4 streptomyce |
| 692 | 7 | 0.9 | 820  | 10 | Q9SXB3 | Q9sxb3 arabidopsis  | 765 | 7 | 0.9 | 1310 | 16 | Q8KR37 | Q8kr37 ralstonia s |
| 693 | 7 | 0.9 | 821  | 5  | Q9VL46 | Q9vl46 drosophila   | 766 | 7 | 0.9 | 1313 | 5  | Q8I353 | Q8i353 plasmodium  |
| 694 | 7 | 0.9 | 830  | 4  | Q8WVZ5 | Q8wvz5 homo sapien  | 767 | 7 | 0.9 | 1374 | 9  | Q9B0D2 | Q9bod2 staphylococ |
| 695 | 7 | 0.9 | 830  | 10 | Q93Z09 | Q93zu9 arabidopsis  | 768 | 7 | 0.9 | 1377 | 3  | Q42695 | Q42695 candida alb |
| 696 | 7 | 0.9 | 833  | 4  | Q9UF54 | Q9uf54 homo sapien  | 769 | 7 | 0.9 | 1398 | 5  | Q9VXU2 | Q9vxu2 drosophila  |
| 697 | 7 | 0.9 | 833  | 12 | Q69403 | Q69403 pseudorabie  | 770 | 7 | 0.9 | 1400 | 5  | Q9N9L6 | Q9n9l6 leishmania  |
| 698 | 7 | 0.9 | 834  | 11 | Q9JL12 | Q9jl12 mus musculus | 771 | 7 | 0.9 | 1439 | 5  | Q8MQC7 | Q8mqc7 caenorhabdi |
| 699 | 7 | 0.9 | 834  | 11 | Q9LW06 | Q9lw06 mus musculus | 772 | 7 | 0.9 | 1463 | 16 | Q9ADP6 | Q9adp6 streptomyce |
| 700 | 7 | 0.9 | 834  | 11 | Q9JHF5 | Q9jhf5 mus musculus | 773 | 7 | 0.9 | 1480 | 5  | Q18138 | Q18138 caenorhabdi |
| 701 | 7 | 0.9 | 837  | 2  | Q933U9 | Q933u9 haemophilus  | 774 | 7 | 0.9 | 1504 | 5  | Q97204 | Q97204 leishmania  |
| 702 | 7 | 0.9 | 837  | 2  | Q933W1 | Q933w1 haemophilus  | 775 | 7 | 0.9 | 1516 | 10 | Q84548 | Q84548 arabidopsis |
| 703 | 7 | 0.9 | 837  | 2  | Q94813 | Q94813 haemophilus  | 776 | 7 | 0.9 | 1536 | 2  | Q48031 | Q48031 haemophilus |
| 704 | 7 | 0.9 | 846  | 10 | Q8H963 | Q8h963 zinnia eleg  | 777 | 7 | 0.9 | 1541 | 5  | Q15837 | Q15837 leishmania  |
| 705 | 7 | 0.9 | 847  | 11 | Q8R3S8 | Q8r3s8 mus musculus | 778 | 7 | 0.9 | 1568 | 5  | Q8I332 | Q8i332 plasmodium  |
| 706 | 7 | 0.9 | 850  | 10 | Q81905 | Q81905 arabidopsis  | 779 | 7 | 0.9 | 1645 | 5  | Q9VIK7 | Q9vik7 drosophila  |
| 707 | 7 | 0.9 | 851  | 10 | Q9S971 | Q9s971 arabidopsis  | 780 | 7 | 0.9 | 1654 | 10 | Q9C5X9 | Q9c5x9 arabidopsis |
| 708 | 7 | 0.9 | 870  | 10 | Q9LQK4 | Q9lqk4 arabidopsis  | 781 | 7 | 0.9 | 1685 | 4  | Q9UEM8 | Q9uem8 homo sapien |
| 709 | 7 | 0.9 | 877  | 5  | Q8XK66 | Q8xk66 drosophila   | 782 | 7 | 0.9 | 1787 | 5  | Q25645 | Q25645 plasmodium  |
| 710 | 7 | 0.9 | 878  | 16 | Q8KCC0 | Q8kcc0 chlorobium   | 783 | 7 | 0.9 | 1789 | 5  | Q9U7Y0 | Q9u7y0 drosophila  |
| 711 | 7 | 0.9 | 880  | 16 | Q8FSG7 | Q8fsg7 corynebacte  | 784 | 7 | 0.9 | 1791 | 5  | Q9U6D4 | Q9ued4 plasmodium  |
| 712 | 7 | 0.9 | 882  | 16 | Q33268 | Q33268 mycobacteri  | 785 | 7 | 0.9 | 1827 | 5  | Q20535 | Q20535 caenorhabdi |
| 713 | 7 | 0.9 | 883  | 10 | Q9M9X1 | Q9m9x1 arabidopsis  | 786 | 7 | 0.9 | 1861 | 5  | Q01401 | Q01401 drosophila  |
| 714 | 7 | 0.9 | 900  | 11 | Q8BY46 | Q8by46 mus musculus | 787 | 7 | 0.9 | 1861 | 5  | Q9VC45 | Q9vc45 drosophila  |
| 715 | 7 | 0.9 | 902  | 12 | Q69095 | Q69095 herpes simp  | 788 | 7 | 0.9 | 1892 | 10 | Q8H6I8 | Q8h6i8 zea mays (m |
| 716 | 7 | 0.9 | 904  | 12 | Q37453 | Q37453 herpes simp  | 789 | 7 | 0.9 | 1929 | 5  | Q9VF02 | Q9vf02 drosophila  |
| 717 | 7 | 0.9 | 904  | 12 | Q89920 | Q89920 herpes simp  | 790 | 7 | 0.9 | 1939 | 5  | Q25662 | Q25662 plasmodium  |
| 718 | 7 | 0.9 | 904  | 12 | Q9LCE3 | Q9lce3 herpes simp  | 791 | 7 | 0.9 | 1962 | 12 | Q91QZ3 | Q91qz3 citrus leaf |
| 719 | 7 | 0.9 | 905  | 5  | Q18260 | Q18260 caenorhabdi  | 792 | 7 | 0.9 | 1966 | 10 | Q9SZW8 | Q9szw8 arabidopsis |
| 720 | 7 | 0.9 | 907  | 16 | Q92KQ9 | Q92kq9 rhizobium m  | 793 | 7 | 0.9 | 2030 | 13 | Q9DDV7 | Q9ddv7 xenopus lae |
| 721 | 7 | 0.9 | 919  | 12 | Q8Q830 | Q8qs30 chimpanzee   | 794 | 7 | 0.9 | 2065 | 13 | Q9PT63 | Q9pt63 xenopus lae |
| 722 | 7 | 0.9 | 925  | 5  | Q18782 | Q18782 caenorhabdi  | 795 | 7 | 0.9 | 2066 | 9  | Q8SDP3 | Q8sdp3 staphylococ |
| 723 | 7 | 0.9 | 927  | 3  | Q92445 | Q92445 pneumocysti  | 796 | 7 | 0.9 | 2066 | 13 | Q9DDV8 | Q9ddv8 xenopus lae |
| 724 | 7 | 0.9 | 929  | 10 | Q82188 | Q82188 arabidopsis  | 797 | 7 | 0.9 | 2066 | 16 | Q8NWK8 | Q8nwk8 staphylococ |
| 725 | 7 | 0.9 | 930  | 2  | Q53173 | Q53173 rhodobacter  | 798 | 7 | 0.9 | 2069 | 16 | Q8NSD9 | Q8nsd9 corynebacte |
| 726 | 7 | 0.9 | 931  | 13 | Q91995 | Q91995 xenopus lae  | 799 | 7 | 0.9 | 2078 | 16 | Q98K31 | Q98k31 rhizobium l |
| 727 | 7 | 0.9 | 935  | 4  | Q96JF8 | Q96jfh homo sapien  | 800 | 7 | 0.9 | 2082 | 16 | Q9S200 | Q9s200 streptomyce |
| 728 | 7 | 0.9 | 949  | 5  | Q97306 | Q97306 plasmodium   | 801 | 7 | 0.9 | 2178 | 2  | Q9KWR3 | Q9kwr3 streptococ  |
| 729 | 7 | 0.9 | 951  | 5  | Q8SLV6 | Q8slv6 encephalito  | 802 | 7 | 0.9 | 2301 | 10 | Q9ATK5 | Q9atk5 chlamydomon |
| 730 | 7 | 0.9 | 956  | 2  | Q93RE7 | Q93re7 bacillus sp  | 803 | 7 | 0.9 | 2303 | 4  | Q95996 | Q95996 homo sapien |
| 731 | 7 | 0.9 | 965  | 5  | Q9T311 | Q9t311 drosophila   | 804 | 7 | 0.9 | 2303 | 16 | Q9ZD91 | Q9zdg1 rickettsia  |
| 732 | 7 | 0.9 | 984  | 10 | Q9LGF1 | Q9lgl1 oryza sativ  | 805 | 7 | 0.9 | 2338 | 16 | Q9LIV8 | Q9liv8 streptomyce |
| 733 | 7 | 0.9 | 985  | 5  | Q95SH0 | Q95sh0 drosophila   | 806 | 7 | 0.9 | 2411 | 16 | Q9RFF2 | Q9rff2 xanthomonas |
| 734 | 7 | 0.9 | 988  | 11 | Q9EQH5 | Q9eqh5 rattus norv  | 807 | 7 | 0.9 | 2491 | 5  | Q97002 | Q97002 leishmania  |
| 735 | 7 | 0.9 | 1006 | 5  | Q8MRK6 | Q8mrk6 drosophila   | 808 | 7 | 0.9 | 2591 | 2  | Q54959 | Q54959 streptomyce |
| 736 | 7 | 0.9 | 1007 | 5  | Q96831 | Q96831 drosophila   | 809 | 7 | 0.9 | 2613 | 5  | Q9GYD1 | Q9gyd1 leishmania  |
| 737 | 7 | 0.9 | 1013 | 4  | Q8ND31 | Q8nd31 homo sapien  | 810 | 7 | 0.9 | 2703 | 5  | Q9VEG7 | Q9veg7 drosophila  |
| 738 | 7 | 0.9 | 1017 | 16 | Q8FRF9 | Q8frf9 corynebacte  | 811 | 7 | 0.9 | 2715 | 5  | Q61603 | Q61603 drosophila  |
| 739 | 7 | 0.9 | 1029 | 10 | Q23015 | Q23015 arabidopsis  | 812 | 7 | 0.9 | 2716 | 5  | Q8IN94 | Q8in94 drosophila  |
| 740 | 7 | 0.9 | 1045 | 4  | Q9NS56 | Q9ns56 homo sapien  | 813 | 7 | 0.9 | 2793 | 16 | Q8X2Q2 | Q8x2q2 escherichia |
| 741 | 7 | 0.9 | 1045 | 4  | Q9UNR9 | Q9unr9 homo sapien  | 814 | 7 | 0.9 | 2806 | 2  | Q9XXA6 | Q9xxa6 escherichia |
| 742 | 7 | 0.9 | 1063 | 10 | Q9XEJ4 | Q9xej4 zea mays (m  | 815 | 7 | 0.9 | 2806 | 9  | Q9TKX6 | Q9tkx6 bacterioph  |
| 743 | 7 | 0.9 | 1072 | 16 | Q92X83 | Q92x83 rhizobium m  | 816 | 7 | 0.9 | 2806 | 9  | Q9XJM1 | Q9xjm1 bacterioph  |
| 744 | 7 | 0.9 | 1089 | 5  | Q26155 | Q26155 plasmodium m | 817 | 7 | 0.9 | 2806 | 16 | Q8X470 | Q8x470 escherichia |
| 745 | 7 | 0.9 | 1093 | 3  | Q8XOR0 | Q8xor0 neurospora   | 818 | 7 | 0.9 | 2812 | 3  | Q74630 | Q74630 schizosacch |
| 746 | 7 | 0.9 | 1103 | 10 | Q9LJY3 | Q9ljy3 arabidopsis  | 819 | 7 | 0.9 | 2812 | 3  | Q8W6K4 | Q8w6k4 sinorhizobi |



956 Q8ahy9 human immun  
 957 Q8ahy5 human immun  
 958 Q984f3 rhizobium l  
 959 Q97x80 sulfobolus  
 960 Q97654 equus caball  
 961 Q97656 ovis aries  
 962 Q9mx50 actus nancy  
 963 Q9mx83 actus nancy  
 964 P89084 human adeno  
 965 Q9yhs1 pseudopleur  
 966 Q99013 pseudopleur  
 967 Q8gi73 salmonella  
 968 Q9zx74 mycobacteri  
 969 Q9xh15 elaeis guin  
 970 Q8k540 mus musculu  
 971 Q9qj86 human immun  
 972 Q9qj85 human immun  
 973 Q9qj84 human immun  
 974 Q9qj83 human immun  
 975 Q9qj82 human immun  
 976 Q9qj81 human immun  
 977 Q8pf19 xanthomonas  
 978 Q8p3m7 xanthomonas  
 979 O07503 borrelia ga  
 980 Q9ayp3 oryza sativ  
 981 Q8jym5 bovine rota  
 982 Q8jym4 bovine rota  
 983 Q8jym3 bovine rota  
 984 Q8jym2 bovine rota  
 985 Q8jym1 bovine rota  
 986 Q8jym7 bovine rota  
 987 Q8jym6 bovine rota  
 988 Q8jym5 bovine rota  
 989 Q8jym4 bovine rota  
 990 Q8jym3 bovine rota  
 991 Q8jym2 bovine rota  
 992 Q8jym1 bovine rota  
 993 Q8jym7 bovine rota  
 994 Q8jym6 bovine rota  
 995 Q8jym5 bovine rota  
 996 Q8jym4 bovine rota  
 997 Q8jym3 bovine rota  
 998 Q8jym2 bovine rota  
 999 Q8jym1 bovine rota  
 1000 Q8jym7 bovine rota

## ALIGNMENTS

## RESULT 1

Q9JYW4  
 ID Q9JYW4 PRELIMINARY; PRT; 761 AA.  
 AC Q9JYW4  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE Pilus secretin.  
 GN Pilus OR NMA0650.  
 OS Neisseria meningitidis (serogroup A).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=65699;  
 RN [1]\_SEQUENCE FROM N.A.  
 RC STRAIN=22491 / Serogroup A / Serotype 4b;  
 RX MEDLINE=20222556; PubMed=10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 RA Davies S.R., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,  
 RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,  
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 RA Whitehead S., Spratt B.G., Barrall B.G.;  
 RT "Complete DNA sequence of a serogroup A strain of Neisseria  
 meningitidis 22491";  
 RL Nature 404:502-506(2000).  
 DR EMBL; AL162753; CAB83938.1; -;  
 DR InterPro; IPR001775; Bac\_GSPD.  
 DR InterPro; IPR004846; GSPII/IITprotein.  
 DR InterPro; IPR004845; GSPII/proteinC.  
 DR InterPro; IPR005644; NOLW-like.  
 DR Pfam; PF00263; GSPII\_III; 1.  
 DR Pfam; PF03958; GSPII\_III\_N; 1.

DR PRINTS; PRO0811; BCTERIALGSPD.  
 DR PROSITE; PS00875; TASP\_D; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 761 AA; 81786 MW; F551769291E07BD5 CRC64;  
 Query Match 71.3%; Score 548; DB 16; Length 761;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 222 PAKQTNIDPRKQKAGIIEAALGAFAGQPDISOQHDHIIIVTLKQHTLPTTLQRLSDVAD 281  
 DB 214 PAKQTNIDPRKQKAGIIEAALGAFAGQPDISOQHDHIIIVTLKQHTLPTTLQRLSDVAD 273  
 QY 282 FXTPTQKVTKLKLNNDTOLIIITAGNWLNVKSAAPGYFTFOVLPKKQNLSEGGVNNAPK 341  
 DB 274 FXTPTQKVTKLKLNNDTOLIIITAGNWLNVKSAAPGYFTFOVLPKKQNLSEGGVNNAPK 333  
 QY 342 TPTGKKSISLDFQDVEIRIILQILAKESGNIVASDSVNGKMTLSLKDVPDQALDLVMOA 401  
 DB 334 TPTGKKSISLDFQDVEIRIILQILAKESGNIVASDSVNGKMTLSLKDVPDQALDLVMOA 393  
 QY 402 RNLDMRQOQNI VNIAPRDELLAKDKAFQAEKDIALGALYSONFQKYNVEEFSILR 461  
 DB 394 RNLDMRQOQNI VNIAPRDELLAKDKAFQAEKDIALGALYSONFQKYNVEEFSILR 453  
 QY 462 LDNADTTGNRNLTVSGRGSVLIDPATNTLIIVTDRSVIEKFRKLIIDELDPVPAQVMIEAR 521  
 DB 454 LDNADTTGNRNLTVSGRGSVLIDPATNTLIIVTDRSVIEKFRKLIIDELDPVPAQVMIEAR 513  
 QY 522 IVEADGFSRDLGVKFGATGKKLNDTSAGFGWVNSGFGGDDKGAETKINLPITAAAN 581  
 DB 514 IVEADGFSRDLGVKFGATGKKLNDTSAGFGWVNSGFGGDDKGAETKINLPITAAAN 573  
 QY 582 SISLVRAISSGALNLELSASELSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANG 641  
 DB 574 SISLVRAISSGALNLELSASELSKTKTLANPRVLTQNRKEAKIESGYEIPFTVTSIANG 633  
 QY 642 GSSTNTEKKAIVGLTVTPNITPDGQIIMTVKINKDSPACASGNOTILCISTKNLTQA 701  
 DB 634 GSSTNTEKKAIVGLTVTPNITPDGQIIMTVKINKDSPACASGNOTILCISTKNLTQA 693  
 QY 702 MVENGSTLIVGGIYBEDNGNTLTKVPLLGDIPIVGNLTKRGKTDRELLIFITPRIMG 761  
 DB 694 MVENGSTLIVGGIYBEDNGNTLTKVPLLGDIPIVGNLTKRGKTDRELLIFITPRIMG 753  
 QY 762 TAGNSLRY 769  
 DB 754 TAGNSLRY 761  
 RESULT 2  
 Q50972  
 ID Q50972 PRELIMINARY; PRT; 720 AA.  
 AC Q50972  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE PilQ.  
 GN PilQ.  
 OS Neisseria gonorrhoeae  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=485;  
 RN [1]\_SEQUENCE FROM N.A.  
 RC STRAIN=MS11;  
 RX MEDLINE=96422484; PubMed=8825101;  
 RA Drake S.L., Komey M.;  
 RT "The product of the pilQ gene is essential for the biogenesis of type  
 IV pili in Neisseria gonorrhoeae";  
 RL Mol. Microbiol. 18:975-986(1995).  
 DR EMBL; U40596; AAC43603.1; -;  
 DR InterPro; IPR001775; Bac\_GSPD.

```

DR InterPro: IPR004846; GSP11/IIIProtein.
DR InterPro: IPR005644; NOLW-like.
DR Pfam: PF00263; GSP11_III; 1.
DR Pfam: PF03958; GSP11_III_N; 1.
DR PRINTS: PR00811; BCTERIALGSPD.
SQ SEQUENCE 720 AA; 77596 MW; A45BE2AD06DEE92B CRC64;

Query Match
Best Local Similarity 100.0%; DB 2; Length 720;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNTKLTIIISGLFVATAAFQTASAGNITDIKVSLSLPNKQKIVKVSFDKKEIVNPTGFTVS 60
DB 1 MNTKLTIIISGLFVATAAFQTASAGNITDIKVSLSLPNKQKIVKVSFDKKEIVNPTGFTVS 60

QY 61 PARIALDFEQTGSMDOQVLEYADPLLSKISAAQNSRARLVNLNPKQGYNTEVRGNKV 120
DB 61 PARIALDFEQTGSMDOQVLEYADPLLSKISAAQNSRARLVNLNPKQGYNTEVRGNKV 120

QY 121 WIFINESDDTVSAPARPAVKAAPAAKQ 150
DB 121 WIFINESDDTVSAPARPAVKAAPAAKQ 150

RESULT 3
Q9ZHF3 PRELIMINARY; PRT; 766 AA.
ID Q9ZHF3
AC Q9ZHF3; 1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Secretin.
GN PilQ.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H44/76;
RX MEDLINE=98367129; PubMed=9701807;
RA Tonium T.; Gaugant D.A.; Dunham S.A.; Kooney M.;
RT "Structure and function of repetitive sequence elements associated
RT with a highly polymorphic domain of the Neisseria meningitidis PilQ
RT protein.";
RL Mol. Microbiol. 29:111-124(1998).
DR EMBL: AF066056; AAC96097.1; .
DR InterPro: IPR001775; Bac GSPD.
DR InterPro: IPR004846; GSP11/IIIProtein.
DR InterPro: IPR005644; NOLW-like.
DR Pfam: PF00263; GSP11_III; 1.
DR Pfam: PF03958; GSP11_III_N; 1.
DR PRINTS: PR00811; BCTERIALGSPD.
SQ SEQUENCE 766 AA; 82087 MW; A538DB3514404D6D CRC64;

Query Match
Best Local Similarity 100.0%; DB 2; Length 766;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNTKLTIIISGLFVATAAFQTASAGNITDIKVSLSLPNKQKIVKVSFDKKEIVNPTGFTVS 60
DB 1 MNTKLTIIISGLFVATAAFQTASAGNITDIKVSLSLPNKQKIVKVSFDKKEIVNPTGFTVS 60

QY 61 PARIALDFEQTGSMDOQVLEYADPLLSKISAAQNSRARLVNLNPKQGYNTEVRGNKV 120
DB 61 PARIALDFEQTGSMDOQVLEYADPLLSKISAAQNSRARLVNLNPKQGYNTEVRGNKV 120

QY 121 WIFINESDDTVSAPARPAVKAAPAAKQ 150
DB 121 WIFINESDDTVSAPARPAVKAAPAAKQ 150

RESULT 4

```

```

Q92N56 PRELIMINARY; PRT; 689 AA.
ID Q92N56
AC Q92N56;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical transmembrane protein SMC02706.
GN R02368 OR SMC02706.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D.; Barloy-Hubler F.; Gouzy J.; Bothe G.; Ampe F.; Batut J.;
RA Boistard P.; Becker A.; Boutry M.; Cadieu E.; Dreano S.; Gloux S.;
RA Godrie T.; Goffeau A.; Kahn D.; Kiss E.; Lelaure V.; Masuy D.;
RA Pohl T.; Portetelle D.; Puehler A.; Purnelle B.; Ramaprasad U.;
RA Renard C.; Thebault P.; Vandenbol M.; Weidner S.; Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
KW Hypothetical protein; Complete proteome.
DR EMBL: AL591790; CAC46947.1; .
SQ SEQUENCE 689 AA; 74875 MW; D49D8BCCD2147A2 CRC64;

Query Match
Best Local Similarity 100.0%; DB 16; Length 689;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 518 IEARIVEAADG 528
DB 103 IEARIVEAADG 113

RESULT 5
Q8XV60 PRELIMINARY; PRT; 714 AA.
ID Q8XV60
AC Q8XV60;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Probable fibrillar type-4 assembly signal peptide protein.
GN PilQ OR RSC2971 OR RS01326.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M.; Genin S.; Attiguenave F.; Gouzy J.; Mangenot S.;
RA Ariat M.; Billault A.; Brottier P.; Camus J.C.; Cattolico L.;
RA Chandler M.; Choisme N.; Claudel-Renard C.; Cunac S.; Demange N.;
RA Gaspin C.; Lavie M.; Moisan A.; Robert C.; Saurin W.; Schiex T.;
RA Guigier P.; Thebault P.; Whalen M.; Wincker P.; Levy M.;
RA Weissenbach J.; Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL: AL646073; CAD16680.1; .
DR InterPro: IPR001775; Bac GSPD.
DR InterPro: IPR004846; GSP11/IIIProtein.
DR InterPro: IPR004845; GSP11proteinc.
DR InterPro: IPR005644; NOLW-like.
DR InterPro: IPR003522; SecIII_OMPG.
DR Pfam: PF00263; GSP11_III; 1.
DR Pfam: PF03958; GSP11_III_N; 1.
DR PRINTS: PR00811; BCTERIALGSPD.
DR PRINTS: PR01337; TYPE3OMGPROT.
DR PROSITE: PS00875; T2SP_D; 1.

```

KW Complete proteome.  
SQ SEQUENCE 714 AA; 76611 MW; 94AE6FB5F8E5A070 CRC64;  
Query Match 1.4%; Score 11; DB 16; Length 714;  
Best Local Similarity 100.0%; Pred. No. 0.075;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 515 QVMIEARIVEA 525  
DB 463 QVMIEARIVEA 473  
|||||

RESULT 6  
Q9AM59 ID Q9AM59 PRELIMINARY; PRT; 723 AA.  
AC Q9AM59; 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Putative outer membrane protein ComQ.  
GN COMQ.  
OS Acinetobacter sp. BD413.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Moraxellaceae; Acinetobacter.  
OX NCBI\_TaxID=104611;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Rosenblauter C., Averhoff B.;  
RT "Genes essential for the natural transformation process in  
RL Acinetobacter sp. BD413.";  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF329876; AAK00351.1; -  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSPiI/IIprotein.  
DR InterPro; IPR005644; NOLW-like.  
DR Pfam; PF00263; GSPiI\_III; 1.  
DR Pfam; PF03958; GSPiI\_III; 1.  
DR PRINTS; PR00811; BCTERIALGSPD.  
DR PROSITE; PS00875; T2SP\_D; 1.  
SQ SEQUENCE 723 AA; 78166 MW; E09AC8369907DAB9 CRC64;

Query Match 1.4%; Score 11; DB 2; Length 723;  
Best Local Similarity 100.0%; Pred. No. 0.076;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 386 LKQVFWQALD 396  
DB 317 LKQVFWQALD 327  
|||||

RESULT 7  
Q32566 ID Q32566 PRELIMINARY; PRT; 585 AA.  
AC Q32566; 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE ETPD protein.  
GN ETPD.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schmidt H., Henkel B., Karch H.;  
RT "A gene cluster closely related to type II secretion pathway operons  
RT of gram-negative bacteria is located on the large plasmid of  
RT enterohemorrhagic Escherichia coli O157 strains.";  
RL FEMS Microbiol. Lett. 148:265-272 (1997).

[2]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=98290540; PubMed=9628576;  
RX Makino K., Ishii K., Yasunaga T., Hattori M., Yokoyama K.,  
RA Yatsudo H.C., Kubota Y., Yamachi Y., Iida T., Yamamoto K., Honda T.,  
RA Han C., Ohtsubo A., Kasamatsu M., Hayashi T., Kuhara S.,  
RA Shinagawa H.;  
RT "Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an  
RT enterohemorrhagic Escherichia coli O157:H7 derived from Sakai  
RT outbreak.";  
RL DNA Res. 5:1-9 (1998).  
DR EMBL; Y09824; CAA70955.1; -  
DR EMBL; AB011549; BAA31759.1; -  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSPiI/IIprotein.  
DR InterPro; IPR005644; NOLW-like.  
DR Pfam; PF00263; GSPiI\_III; 1.  
DR Pfam; PF03958; GSPiI\_III; 3.  
DR PRINTS; PR00811; BCTERIALGSPD.  
DR PROSITE; PS00875; T2SP\_D; 1.  
KW Plasmid.  
SQ SEQUENCE 585 AA; 63614 MW; 62AE17CAD87A24FC CRC64;  
Query Match 1.3%; Score 10; DB 2; Length 585;  
Best Local Similarity 100.0%; Pred. No. 0.68;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDIPIV 734  
DB 485 KVPLLGDIPIV 494  
|||||

RESULT 8  
Q47423 ID Q47423 PRELIMINARY; PRT; 596 AA.  
AC Q47423; 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE PULD-like protein.  
GN PULD-LIKE.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=EDL 933;  
RA Schmidt H.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X86372; CAA60131.1; -  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSPiI/IIprotein.  
DR InterPro; IPR005644; NOLW-like.  
DR Pfam; PF00263; GSPiI\_III; 1.  
DR Pfam; PF03958; GSPiI\_III; 3.  
DR PRINTS; PR00811; BCTERIALGSPD.  
KW Plasmid.  
SQ SEQUENCE 596 AA; 64912 MW; 3354901B42F981DF CRC64;  
Query Match 1.3%; Score 10; DB 2; Length 596;  
Best Local Similarity 100.0%; Pred. No. 0.7;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDIPIV 734  
DB 496 KVPLLGDIPIV 505  
|||||

RESULT 9  
Q9ZGU0

```

ID Q9ZGU0 PRELIMINARY; PRT; 642 AA.
AC Q9ZGU0;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Type II secretion protein.
GN EIPD.
OS Escherichia coli O157:H7.
OG Plasmid pO157.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EDL933;
RX MEDLINE=98391744; PubMed=9722640;
RA Burland V., Shao Y., Perna N.T., Plunkett G., Sofia H.J.,
RA Blattner F.R.;
RT "The complete DNA sequence and analysis of the large virulence plasmid
of Escherichia coli O157:H7.";
RL Nucleic Acids Res. 26:4196-4204 (1998).
DR EMBL; AF074613; AAC70101.1; -.
DR InterPro; IPR001775; BacGSPD.
DR InterPro; IPR004846; GSPII/IIprotein.
DR InterPro; IPR004845; GSPII/IIproteinC.
DR InterPro; IPR005644; NOLW-like.
DR Pfam; PF00263; GSPII_III; 1.
DR Pfam; PF03958; GSPII_III_N; 3.
DR PRINTS; PR00811; BCTERIALGSPD.
DR PROSITE; PS00875; T2SP_D; 1.
KW Plasmid.
SQ SEQUENCE 642 AA; 69911 MW; FB5E74CC1DC2B4DC CRC64;

Query Match 1.3%; Score 10; DB 2; Length 642;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 725 KVPLLGDPV 734
Db 542 KVPLLGDPV 551
|||||
|

RESULT 10
ID Q8DDT0 PRELIMINARY; PRT; 673 AA.
AC Q8DDT0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Type II secretory pathway, component EpsD.
GN V10877.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016799; AA009380.1; -.
KW Complete proteome.
SQ SEQUENCE 673 AA; 73491 MW; 81E7D70F0D0C025 CRC64;

Query Match 1.3%; Score 10; DB 16; Length 673;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 725 KVPLLGDPV 734
Db 575 KVPLLGDPV 584
|||||
|


```

```

RESULT 11
Q8EK21 PRELIMINARY; PRT; 684 AA.
AC Q8EK21;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Type IV pilus biogenesis protein PilQ.
GN SO0285.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=2297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Kfoury H., Gill J., Uterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123 (2002).
DR EMBL; AE015476; AAN53370.1; -.
DR TIGR; SO0285; -.
KW Complete proteome.
SQ SEQUENCE 684 AA; 74857 MW; DAEEFF7B9F7408 CRC64;

Query Match 1.3%; Score 10; DB 16; Length 684;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 388 DVPWDQALDL 397
Db 314 DVPWDQALDL 323
|||||
|

RESULT 12
Q9ZF86 PRELIMINARY; PRT; 750 AA.
AC Q9ZF86;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE General secretory pathway protein D.
GN GSPD.
OS Burkholderia pseudomallei (Pseudomonas pseudomallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=28450;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1026b;
RX MEDLINE=93350433; PubMed=10419967;
RA DeShazer D., Brett P.J., Burtnick M.N., Woods D.E.;
RT "Molecular characterization of genetic loci required for secretion of
exoproducts in Burkholderia pseudomallei.";
RL J. Bacteriol. 181:4661-4664 (1999).
DR EMBL; AF10185; AAD05173.1; -.
DR InterPro; IPR001775; BacGSPD.
DR InterPro; IPR004846; GSPII/IIprotein.
DR InterPro; IPR005644; NOLW-like.
DR Pfam; PF00263; GSPII_III; 1.
DR Pfam; PF03958; GSPII_III_N; 3.
DR PRINTS; PR00811; BCTERIALGSPD.
SQ SEQUENCE 750 AA; 77509 MW; 7D5852D07E53F217 CRC64;


```



Query Match 1.3%; Score 10; DB 2; Length 750;  
Best Local Similarity 100.0%; Pred. No. 0.86;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 724 TKVPLIGDIP 733  
DB 623 TKVPLIGDIP 632

RESULT 13

Q92FG1 PRELIMINARY; PRT; 901 AA.  
AC Q92FG1  
DT 01-MAY-1999 (T-EMBLrel. 10, Created)  
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)  
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
DE Pfam: PF01406; trna-synt\_le; 1.  
DE PRINTS: PR00983; TRNASYNTHCYS.  
DE TIGRFAMs: TIGR00435; cysS; 1.  
SQ SEQUENCE 908 AA; 105217 MW; 0B3BC4B4F9772B35 CRC64;

Query Match 1.3%; Score 10; DB 5; Length 908;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 DTRSVIEKFR 503  
DB 673 DTRSVIEKFR 682

RESULT 15

Q76618 PRELIMINARY; PRT; 909 AA.  
AC Q76618  
DT 01-NOV-1998 (T-EMBLrel. 08, Created)  
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)  
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
DE Hypothetical 105.3 kDa protein.  
DE Y23HSA.7.  
GN Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RT Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Dempsey S., Le T.T.;  
RT "The sequence of C. elegans cosmid Y23HSA.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RT "Direct Submission";  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF077541; AAC64633.1; -.  
DR WormPep: Y23HSA.7a; CE18363.  
DR InterPro: IPR002308; Cys\_trna-synt\_la.  
DR Pfam: PF01406; trna-synt\_le; 1.  
DR PRINTS: PR00983; TRNASYNTHCYS.  
DR TIGRFAMs: TIGR00435; cysS; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 909 AA; 105316 MW; FDF4E96133864DAC CRC64;

Query Match 1.3%; Score 10; DB 5; Length 909;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 389 VPWDQALDLV 398  
DB 523 VPWDQALDLV 532

RESULT 14

Q96516 PRELIMINARY; PRT; 908 AA.  
AC Q96516  
DT 01-DEC-2001 (T-EMBLrel. 19, Created)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
DE Hypothetical protein Y23HSA.7b.  
DE Y23HSA.7.  
GN Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RT Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.

QY 494 DTRSVIEKFR 503  
 |||||  
 DB 674 DTRSVIEKFR 683

## RESULT 16

Q8RCK3 PRELIMINARY; PRT; 214 AA.  
 AC Q8RCK3;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical protein TTE0424.  
 GN TTE0424.  
 OS Thermoanaerobacter tengcongensis.  
 OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;  
 OC Thermoanaerobacteriaceae; Thermoanaerobacter.  
 OX NCBI\_TaxID=119072;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MB4 / JCM 11007;  
 EX MEDLINE=21992816; PubMed=1197336;  
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;  
 RT "A complete sequence of T. tengcongensis genome.";  
 RL Genome Res. 12:689-700(2002).  
 DR EMBL: AE013014; AAM23708.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 214 AA; 24385 MW; 71A2A182FAEBB316 CRC64;

Query Match 1.2%; Score 9; DB 16; Length 214;  
 Best Local Similarity 100.0%; Pred. No. 3;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 KIVKVSFDK 48  
 |||||  
 DB 74 KIVKVSFDK 82

## RESULT 17

O15562 PRELIMINARY; PRT; 248 AA.  
 AC O15562;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Glutamyl-tRNA synthetase (Fragment).  
 OS Nosema locustae.  
 OC Eukaryota; Fungi; Microsporidia; Nosematidae; Nosema.  
 OX NCBI\_TaxID=33696;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 30860;  
 RA Brown J.R., Doolittle W.F.;  
 RT "Gene descent, duplication, and horizontal transfer in the evolution  
 of glutamyl-tRNA and glutamyl-tRNA synthetases.";  
 RL J. Mol. Evol. 0:0-0(1997)  
 DR EMBL: AF005490; AAB62549.1; -;  
 DR HSP; P00962; IGTR.  
 DR InterPro: IPR000924; Glu\_tRNA-synt\_ic.  
 DR Pfam: PF00749; tRNA-synt\_ic; 1.  
 DR PRINTS; PR00987; TRNASYNTHGLU.  
 KW Aminoacyl-tRNA synthetase.  
 FT NON\_TER 1  
 FT NON\_TER 248 248  
 SQ SEQUENCE 248 AA; 28778 MW; 9FACB858A1D9EC04 CRC64;

Query Match 1.2%; Score 9; DB 5; Length 248;  
 Best Local Similarity 100.0%; Pred. No. 3.4;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 703 VENGTLIV 711

DB 19 VENGTLIV 27  
 |||||

## RESULT 18

Q985G3 PRELIMINARY; PRT; 379 AA.  
 AC Q985G3;  
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Cytochrome c-type biogenesis protein, CycH.  
 GN MLR7688.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 OX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NAFR303099;  
 EX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 Mesorhizobium loti.";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL: AF003012; BAB54099.1; -;  
 DR InterPro: IPR001440; TPR.  
 DR Pfam: PF00515; TPR; 3.  
 KW Complete proteome.  
 SQ SEQUENCE 379 AA; 39556 MW; 4E350892BE70A6D6 CRC64;

Query Match 1.2%; Score 9; DB 16; Length 379;  
 Best Local Similarity 100.0%; Pred. No. 5;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 ILRLDNADT 467  
 |||||  
 DB 72 ILRLDNADT 80

## RESULT 19

Q9JRS7 PRELIMINARY; PRT; 460 AA.  
 AC Q9JRS7;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE OrfY, OrfX, flp, OrfA, OrfB, OrfC, OrfD, OrfE genes, complete and  
 DE partial cds.  
 OS Actinobacillus actinomycetemcomitans (Haemophilus  
 actinomycetemcomitans)  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Actinobacillus.  
 OX NCBI\_TaxID=714;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=98285392; PubMed=9623911;  
 RA Inoue T., Tanimoto I., Ohta H., Kato K., Murayama Y., Fukui K.;  
 RT "Molecular characterization of low-molecular-weight component protein,  
 Flp, in Actinobacillus actinomycetemcomitans fimbriae.";  
 RL Microbiol. Immunol. 42:253-258(1998).  
 DR EMBL: AB005741; BAA96101.1; -;  
 DR InterPro: IPR001775; Bac\_GSPD.  
 DR InterPro: IPR004846; GSPII/IIIprotein.  
 DR Pfam: PF00263; GSPII\_III; 1.  
 DR PRINTS; PR00811; BCTERIALGSPD.  
 SQ SEQUENCE 460 AA; 50113 MW; 092D497E49B63525 CRC64;

Query Match 1.2%; Score 9; DB 2; Length 460;

```

Best Local Similarity 100.0%; Pred. No. 6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 725 KVPLLGDP 733
Db 377 KVPLLGDP 385

RESULT 20
Q9X6J1 PRELIMINARY; PRT; 460 AA.
ID O9X6J1
AC O9X6J1
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Rough colony protein A.
GN RCPA.
OS Actinobacillus actinomycetemcomitans (Haemophilus
actinomycetemcomitans).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Actinobacillus.
OX NCBI_TaxID=714;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=283;
RX MEDLINE=99270950; PubMed=10338497;
RA Haase E.M., Zmuda J.L., Scannapieco F.A.;
RT "Identification and molecular analysis of rough-colony-specific outer
membrane proteins of Actinobacillus actinomycetemcomitans.";
RL Infect. Immun. 67:2501-2508(1999).
DR EMBL; AF139249; AAD29695.1; -.
DR InterPro; IPR001775; Bac GSPD.
DR Pfam; PF00263; GSP11 III; 1.
DR PRINTS; PRO0811; BCTERIALGSPD.
SQ SEQUENCE 460 AA; 50172 MW; 4B7F837AD104CC19 CRC64;

Query Match 1.2%; Score 9; DB 2; Length 460;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 725 KVPLLGDP 733
Db 377 KVPLLGDP 385

RESULT 21
Q8GD00 PRELIMINARY; PRT; 460 AA.
ID Q8GD00
AC Q8GD00
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE RCPA.
GN RCPA.
OS Actinobacillus actinomycetemcomitans (Haemophilus
actinomycetemcomitans).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Actinobacillus.
OX NCBI_TaxID=714;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CUI1000;
RX MEDLINE=21438116; PubMed=11553455;
RA Kachlany S.C., Planet P.J., Desalle R., Fine D.H., Figurski D.H.;
RT "Genes for tight adherence of Actinobacillus actinomycetemcomitans:
from plaque to plague to pond scum.";
RL Trends Microbiol. 9:429-437(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CUI1000;
RA Planet P.J., Kachlany S.C., Fine D.H., Desalle R., Figurski D.H.;
RT "The Widespread Colonization Island of Actinobacillus

```

```

actinomycetemcomitans.";
RT Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY157714; AAN75208.1; -.
SQ SEQUENCE 460 AA; 50212 MW; 4B937C976C08479A CRC64;

Query Match 1.2%; Score 9; DB 2; Length 460;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 725 KVPLLGDP 733
Db 377 KVPLLGDP 385

RESULT 22
Q9CMH4 PRELIMINARY; PRT; 470 AA.
ID Q9CMH4
AC Q9CMH4
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE RCPA OR PM0852.
GN RCPA OR PM0852.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AB006123; AAK02936.1; -.
DR InterPro; IPR001775; Bac GSPD.
DR Pfam; PF004846; GSP11/III protein.
DR PRINTS; PRO0811; BCTERIALGSPD.
KW Complete proteome.
SQ SEQUENCE 470 AA; 51116 MW; 30E1D6D63616BF46 CRC64;

Query Match 1.2%; Score 9; DB 16; Length 470;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 725 KVPLLGDP 733
Db 387 KVPLLGDP 395

RESULT 23
Q9AJ21 PRELIMINARY; PRT; 512 AA.
ID Q9AJ21
AC Q9AJ21
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE EscC (EscC protein).
GN ESCC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RDEC-1;
RX MEDLINE=98254123; PubMed=9593291;
RA Elliott S.J., Wainwright L.A., McDaniel T.K., Jarvis K.G., Deng Y.K.,
RA Lai L.C., McNamara B.P., Donnerberg M.S., Kaper J.B.;
RT "The complete sequence of the locus of enterocyte effacement (LEE)
from enteropathogenic Escherichia coli E2348/69.";
RL Mol. Microbiol. 28:1-4(1998).

```

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=RDSC-1;
RX  MEDLINE=21153569; PubMed=11254564;
RA  Zhu C., Agin T.S., Elliott S.J., Johnson L.A., Thate T.E., Kaper J.B.,
RA  Bodeker E.C.;
RT  "Complete Nucleotide Sequence and Analysis of the Locus of Enterocyte
RT  Effacement from Rabbit Diarrheagenic Escherichia coli RDEC-1.";
RL  Infect. Immun. 69:2107-2115(2001).
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN=83/39;
RA  Tauschek M., Strugnell R.A., Robins-Browne R.M.;
RT  "Characterization of the LEE pathogenicity islands of rabbit
RT  enteropathogenic Escherichia coli.";
RL  Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN  [4]
RP  SEQUENCE FROM N.A.
RC  STRAIN=413/89-1;
RA  Benkel P., Chakraborty T.;
RT  "Genetic organisation and sequence of the LEE II locus in Shiga toxin-
RT  producing Escherichia coli.";
RL  Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AF200363; AAK26708.1; -
DR  EMBL; AF453441; AAL57535.1; -
DR  EMBL; AJ277443; CAC81855.1; -
DR  InterPro; IPR001775; Bac_GSPD.
DR  InterPro; IPR004846; GSP1/II protein.
DR  InterPro; IPR005644; NslW-like.
DR  InterPro; IPR003522; SecIII_ONPG.
DR  Pfam; PF00263; GSP1_III; 1.
DR  Pfam; PF03958; GSP1_III_N; 1.
DR  PRINTS; PR00811; BCTERIALGSPD.
DR  PRINTS; PR01337; TYPE3OMGPROT.
DR  PROSITE; PS00875; T2SP_D; 1.
SQ  SEQUENCE 512 AA; 56573 MW; 03A0F2A42C9FD831 CRC64;

Query Match 1.2%; Score 9; DB 2; Length 512;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 DIPVIGNLF 739
DB 477 DIPVIGNLF 485
|||||
PRELIMINARY; PRT; 607 AA.

RESULT 24
Q93KTL ID Q93KTL PRELIMINARY; PRT; 607 AA.
AC Q93KTL;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE YSCC.
GN YSCC.
OS Versinia enterocolitica.
OG Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Versinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8081;
RX MEDLINE=21295118; PubMed=11402007;
RA Snellings N.J., Popek M., Lindler L.E.;
RT "Complete DNA Sequence of Versinia enterocolitica Serotype 0:8 Low-
RT Calcium-Response Plasmid Reveals a New Virulence Plasmid-Associated
RT Replicon.";
RL Infect. Immun. 69:4627-4638(2001).
DR EMBL; AF336309; AAK69235.1; -
DR InterPro; IPR001775; Bac_GSPD.
DR InterPro; IPR004846; GSP1/II protein.

```

```

DR InterPro; IPR004845; GSP1IIProteinC.
DR InterPro; IPR005644; NslW-like.
DR InterPro; IPR003522; SecIII_OMPG.
DR Pfam; PF00263; GSP1_III; 1.
DR Pfam; PF03958; GSP1_III_N; 2.
DR PRINTS; PR00811; BCTERIALGSPD.
DR PRINTS; PR01337; TYPE3OMGPROT.
DR PROSITE; PS00875; T2SP_D; 1.
KW Plasmid.
SQ SEQUENCE 607 AA; 67222 MW; 3F98FBA639F1777E CRC64;

Query Match 1.2%; Score 9; DB 2; Length 607;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KYPLLGIDP 733
DB 474 KYPLLGIDP 482
|||||
PRELIMINARY; PRT; 607 AA.

RESULT 25
Q56974 ID Q56974 PRELIMINARY; PRT; 607 AA.
AC Q56974;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE KIM5.
GN KIM5 OR YSCC OR YPCD1.52.
OS Versinia pestis.
OG Plasmid pCD1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Versinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92325077; PubMed=1624469;
RA Haddix P.L., Straley S.C.;
RT "Structure and regulation of the Versinia pestis yscBCDEF operon.";
RL J. Bacteriol. 174:4820-4828(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5;
RX MEDLINE=98422474; PubMed=9748454;
RA Hu P., Elliott J., McCreedy P., Skowronski E., Garnes J.,
RA Kobayashi A., Brubaker R.R., Garcia E.;
RT "Structural organization of virulence-associated plasmids of Versinia
RT pestis.";
RL J. Bacteriol. 180:5192-5202(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5;
RX MEDLINE=98422474; PubMed=9748454;
RA Hu P., Elliott J., McCreedy P., Skowronski E., Garnes J.,
RA Kobayashi A., Brubaker R.R., Garcia E.;
RT "Structural organization of virulence-associated plasmids of Versinia
RT pestis.";
RL J. Bacteriol. 180:5192-5202(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis; PLASMID=pCD1;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titchall R.W., Holden M.T.G.,
RA Prentice M.B., Sebaibia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth I., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Versinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL; M83225; AAK27638.1; -
DR EMBL; AF053946; AAC62552.1; -

```

DR EMBL; AF074612; AAC59781.1; -  
DR EMBL; AL117189; CAB54925.1; -  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSPII/IIIprotein.  
DR InterPro; IPR004845; GSPIIproteinC.  
DR InterPro; IPR005644; NoliW-like.  
DR InterPro; IPR003522; SecIII OMPG.  
DR Pfam; PF00263; GSPII\_III; 1.  
DR Pfam; PF03958; GSPII\_III; 1.  
DR PRINTS; PR00811; BCTERIALGSPD.  
DR PRINTS; PR01337; TYPE3OMGPRD.  
DR PROSITE; PS00875; T2SP.D; 1.  
KW Plasmid; Complete proteome.  
SQ SEQUENCE 607 AA; 67344 MW; A8DA2AF1BFD8083 CRC64;  
  
Query Match 1.2%; Score 9; DB 16; Length 607;  
Best Local Similarity 100.0%; Pred. No. 7.7; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 725 KVPLLGDIIP 733  
Db 474 KVPLLGDIIP 482  
  
RESULT 26  
Q8VRNO PRELIMINARY; PRT; 616 AA.  
AC Q8VRNO;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Hypothetical type II secretion protein GSPD.  
GN GSPD.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]\_TaxID=562;  
RP SEQUENCE FROM N.A.  
RC STRAIN=83/39;  
RA Tauschek M., Gortell R.J., Strugnell R.A., Robins-Browne R.M.;  
RT "Identification of a type II protein secretory pathway required for  
the secretion of heat-labile enterotoxin by enterotoxigenic  
RT Escherichia coli.";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF426313; AAL60184.1; -  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSPII/IIIprotein.  
DR InterPro; IPR005644; NoliW-like.  
DR Pfam; PF00263; GSPII\_III; 1.  
DR Pfam; PF03958; GSPII\_III; 3.  
DR PRINTS; PR00811; BCTERIALGSPD.  
SQ SEQUENCE 616 AA; 66319 MW; 7069455A3F19A8B4 CRC64;  
  
Query Match 1.2%; Score 9; DB 2; Length 616;  
Best Local Similarity 100.0%; Pred. No. 7.8; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 725 KVPLLGDIIP 733  
Db 520 KVPLLGDIIP 528  
  
RESULT 27  
Q8VPC8 PRELIMINARY; PRT; 616 AA.  
AC Q8VPC8;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Hypothetical type II secretion protein.  
GN GSPD.  
OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]\_TaxID=562;  
RP SEQUENCE FROM N.A.  
RC STRAIN=H10407;  
RA Tauschek M., Strugnell R.A., Robins-Browne R.M.;  
RT "Identification of a type II secretory pathway required for secretion  
of heat-labile enterotoxin by Enterotoxigenic Escherichia coli.";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY056599; AAL10693.1; -  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSPII/IIIprotein.  
DR InterPro; IPR005644; NoliW-like.  
DR Pfam; PF00263; GSPII\_III; 1.  
DR Pfam; PF03958; GSPII\_III; 3.  
DR PRINTS; PR00811; BCTERIALGSPD.  
SQ SEQUENCE 616 AA; 66353 MW; 7069455A3F19A654 CRC64;  
  
Query Match 1.2%; Score 9; DB 2; Length 616;  
Best Local Similarity 100.0%; Pred. No. 7.8; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 725 KVPLLGDIIP 733  
Db 520 KVPLLGDIIP 528  
  
RESULT 28  
Q9ZFY0 PRELIMINARY; PRT; 649 AA.  
ID Q9ZFY0  
AC Q9ZFY0;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Outer membrane secretion protein Q.  
GN XCPQ.  
OS Pseudomonas alcaligenes.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=43263;  
RN [1]\_TaxID=43263;  
RP SEQUENCE FROM N.A.  
RC STRAIN=M-1;  
RX MEDLINE=99040991; PubMed=9823657;  
RA Garritse G., Ure R., Bizoullier F., Quax W.J.;  
RT "The phenotypic enhancement method identifies the xcp outer membrane  
RT secretion machinery from Pseudomonas alcaligenes as a bottleneck for  
lipase production.";  
RL J. Biotechnol. 64:23-38(1998).  
DR EMBL; AF092918; AAC83352.1; -  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSPII/IIIprotein.  
DR InterPro; IPR004845; GSPIIproteinC.  
DR InterPro; IPR005644; NoliW-like.  
DR Pfam; PF00263; GSPII\_III; 1.  
DR Pfam; PF03958; GSPII\_III; 3.  
DR PRINTS; PR00811; BCTERIALGSPD.  
DR PROSITE; PS00875; T2SP.D; 1.  
SQ SEQUENCE 649 AA; 65612 MW; 2DE50A042C19B684 CRC64;  
  
Query Match 1.2%; Score 9; DB 2; Length 649;  
Best Local Similarity 100.0%; Pred. No. 8.2; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 725 KVPLLGDIIP 733  
Db 563 KVPLLGDIIP 571  
  
RESULT 29  
Q8CVN5 PRELIMINARY; PRT; 654 AA.  
ID Q8CVN5

```

AC Q8CVN5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DE Hypothetical protein AQ_585.
DN
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Probable general secretion pathway protein D precursor.
DN
GN YHEF OR C4096.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
[1]
RN
SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=2338234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roessch P.,
Rasko D., Buckles E.L., Liou S.-A., Boutin A., Hackett J., Stroud D.,
Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR ENBL; AE016767; AN82534.1; -.
KW Complete proteome.
SQ SEQUENCE 654 AA; 71147 MW; 2C9822E0B39EDF60 CRC64;

Query Match 1.2%; Score 9; DB 16; Length 654;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDIP 733
Db |||||
563 KVPLLGDIP 571

RESULT 30
Q8GBE6 PRELIMINARY; PRT; 658 AA.
AC Q8GBE6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DN
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE YtsLd protein.
DN
GN YtsLd.
OS Yersinia enterocolitica (type O:8).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=34054;
[1]
RN
SEQUENCE FROM N.A.
RC STRAIN=WA-314;
RA Iwobi A., Rakin A., Heesemann J.;
RT "Representational difference analysis reveals a novel type II
secretion cluster unique to highly pathogenic Yersinia enterocolitica
strains."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL; AJ344214; CAC83029.1; -.
SQ SEQUENCE 658 AA; 71592 MW; F91539A6D64230B3 CRC64;

Query Match 1.2%; Score 9; DB 2; Length 658;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDIP 733
Db |||||
561 KVPLLGDIP 569

RESULT 31
O66850 PRELIMINARY; PRT; 705 AA.
AC O66850;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DN
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)

```

```

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein AQ_585.
DN
GN AQ_585.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
[1]
RN
SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aolicus."
RL Nature 392:353-358 (1998).
DR ENBL; AE000697; AAC06820.1; -.
DR InterPro; IPR001775; Bac GSPD.
DR InterPro; IPR004846; GSPII/IIIprotein.
DR Pfam; PF00263; GSPII_III; 1.
DR PRINTS; PR00811; BCTERIALGSPD.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 705 AA; 80771 MW; 2AB9870C1EDD61DF CRC64;

Query Match 1.2%; Score 9; DB 16; Length 705;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 749 RELLIFITP 757
Db |||||
691 RELLIFITP 699

RESULT 32
Q9FIQ1 PRELIMINARY; PRT; 783 AA.
AC Q9FIQ1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DN
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE GspD.
DN
GN GspD.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=292;
[1]
RN
SEQUENCE FROM N.A.
RC STRAIN=KF1;
RA Abe M., Kimoto M., Nakazawa T.;
RT "Molecular organization of the gsp gene cluster in Burkholderia
cepacia."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR ENBL; AB050004; BAB18788.1; -.
DR InterPro; IPR001775; Bac GSPD.
DR InterPro; IPR004846; GSPII/IIIprotein.
DR Pfam; PF00263; GSPII_III; 1.
DR Pfam; PF03958; GSPII_III_N; 3.
DR PRINTS; PR00811; BCTERIALGSPD.
SQ SEQUENCE 783 AA; 81399 MW; 8C7E4CEBF215F662 CRC64;

Query Match 1.2%; Score 9; DB 2; Length 783;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KVPLLGDIP 733
Db |||||
640 KVPLLGDIP 648

RESULT 33
Q03995

```

```

ID Q03995 PRELIMINARY; PRT; 142 AA.
AC Q03995;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE Merozoite surface antigen 1 (fragment).
GN MSA-1.
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5825;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92018030; PubMed=1922203;
RA Deleersnijder W., Hendrix D., Hamers R.;
RT "Analysis of MSA-1 diversity in Plasmodium chabaudi chabaudi strains."
RL Mol. Biochem. Parasitol. 46:315-318(1991).
RC STRAIN=AS;
CC -!- SUBCELLULAR LOCATION: COULD BE ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -!- MISCELLANEOUS: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83K, 42K & 19K ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES, THE MATURATION TAKE PLACE DURING SCHIZONT.
CC EMBL; M61204; AA229514.1; -.
DR EMBL; M61205; AA229452.1; -.
DR InterPro; IPR002965; P_Fich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
KW GPI-anchor; Glycoprotein; Malaria; Merozoite; Polyprotein; Repeat;
KW Signal; Transmembrane.
FT NON_TER 1
FT NON_TER 142
FT NON_TER 142
SQ SEQUENCE 142 AA; 14415 MW; 19B22175D951FB8B CRC64;

Query Match 1.0%; Score 8; DB 5; Length 142;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPAK 148
Db 120 AAPAAPAK 127

RESULT 34
Q04000
ID Q04000 PRELIMINARY; PRT; 142 AA.
AC Q04000;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE Merozoite surface antigen 1 (fragment).
GN MSA-1.
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5825;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=AS;
RX MEDLINE=92018030; PubMed=1922203;
RA Deleersnijder W., Hendrix D., Hamers R.;
RT "Analysis of MSA-1 diversity in Plasmodium chabaudi chabaudi strains."
RL Mol. Biochem. Parasitol. 46:315-318(1991).
RC STRAIN=AS;
CC -!- SUBCELLULAR LOCATION: COULD BE ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -!- MISCELLANEOUS: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83K, 42K & 19K ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES, THE MATURATION TAKE PLACE DURING SCHIZONT.
CC EMBL; M61204; AA229514.1; -.
DR EMBL; M61205; AA229452.1; -.
DR InterPro; IPR002965; P_Fich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
KW GPI-anchor; Glycoprotein; Malaria; Merozoite; Polyprotein; Repeat;
KW Signal; Transmembrane.
FT NON_TER 1
FT NON_TER 142
FT NON_TER 142
SQ SEQUENCE 142 AA; 14401 MW; 1DE32175DD00F89A CRC64;

Query Match 1.0%; Score 8; DB 5; Length 142;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPAK 148
Db 120 AAPAAPAK 127

RESULT 35
Q04000
ID Q04000 PRELIMINARY; PRT; 149 AA.
AC Q04000;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DE Hypothetical protein.
GN OSJNB0091N21.38.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Kim M.M.,
RA Overton II L.L., Bera J.J., Tsitrin T., Krol M.I., Jarzahi B.B.,
RA Jin S.S., Koo H., Zismann V., Hsiao J., Blunt S., Vanaken S.S.,
RA Uterback T.T., Feldblyum T.V., Yang Q.Q., Haas B.J., Suh B.B.,
RA Peterson J.J., Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 10 BAC OSJNB0091N21 genomic sequence."
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC091122; AA094939.1; -.
DR Gramene; Q8LN85; -.
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZE1.
KW Hypothetical protein.
RN [1]
RP SEQUENCE 149 AA; 15534 MW; 119CB662003755BB CRC64;

Query Match 1.0%; Score 8; DB 10; Length 149;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPAK 148
Db 73 AAPAAPAK 80

RESULT 36
Q062S9
ID Q062S9 PRELIMINARY; PRT; 159 AA.
AC Q062S9;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE Ribosomal protein L35A.
OS Spodoptera frugiperda (Fall armyworm).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;
OC Noctuidae; Amphipyrinae; Spodoptera.
OX NCBI_TaxID=7108;
RN [1]
RP SEQUENCE FROM N.A.
RA Landais I., Ogliastro M., Mita K., Nohata J., Lopez-Ferber M.,
RA Duonot-Cerutti M., Fournier P., Devauchelle G.;
RT "Full-length ribosomal protein sequence from an EST library of Spodoptera frugiperda cells (Sf9)."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF400197; AA092169.1; -.
DR InterPro; IPR001780; Ribosomal L35AE.
DR Pfam; PF01247; Ribosomal L35AE; 1.
DR ProDom; PD012670; Ribosomal L35AE; 1.
DR PROSITE; PS01105; RIBOSOMAL_L35AE; 1.
SQ SEQUENCE 159 AA; 17393 MW; C1763824A65E7BDD CRC64;

Query Match 1.0%; Score 8; DB 5; Length 159;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 KAAPAPAK 147

```

```

DR PROSITE; PS00190; CYTOCHROME_C; 1.
KW Complete proteome.
SQ SEQUENCE 215 AA; 21863 MW; 51A30B23D2D9D96E CRC64;

Query Match
Best Local Similarity 1.0%; Score 8; DB 16; Length 215;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPAK 148
DB 208 AAPAAPAK 215
|||||

RESULT 39
Q8XR79 PRELIMINARY; PRT; 222 AA.
ID Q8XR79; AC Q8XR79;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Probable nitrate/nitrite response regulator transcription regulator
DE protein.
GN NARL OR RSP0980 OR RS02312.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billaut A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Chouane N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Signier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646082; CAD18131.1; -.
DR HSSP; P06143; 1D4Z.
DR InterPro; IPR000792; HTH LuxR.
DR InterPro; IPR001789; Response_reg.
DR Pfam; PF00196; GcrE; 1.
DR Pfam; PF00072; response_reg; 1.
DR ProDom; PD000307; HTH_LuxR; 1.
DR ProDom; PD000039; Response_reg; 1.
DR SMART; SM00421; HTH_LuxR; 1.
DR SMART; SM00448; REC; 1.
DR PROSITE; PS00622; HTH_LUXR_FAMILY; 1.
DR PROSITE; PS01110; RESPONSE_REGULATORY; 1.
KW Plasmid; Complete proteome.
SQ SEQUENCE 222 AA; 23957 MW; ADE2035D88D7E91A CRC64;

Query Match
Best Local Similarity 1.0%; Score 8; DB 16; Length 222;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 KAAPAAPA 147
DB 143 KAAPAAPA 150
|||||

RESULT 40
Q9JXY0 PRELIMINARY; PRT; 231 AA.
ID Q9JXY0; AC Q9JXY0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Mannose-1-phosphate guanylttransferase-related protein.
DE NMB1841.
GN NMB1841.

```

```

DB 26 KAAPAAPA 33
|||||

RESULT 37
Q94GX7 PRELIMINARY; PRT; 184 AA.
ID Q94GX7; AC Q94GX7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 19.7 kDa protein.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoideae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cnv. Nipponbare;
RA Buell C.R., Yuan Q., Guyang S., Liu J., Moffat K.S., Hill J.N.,
RA Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
RA Tsitrin T., Riggs F., Hsiao J., Zismann V., Blunt S., Pai G.,
RA VanAken S.E., Uterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
RA Salzberg S.L., White O., Fraser C.M.;
RT "Oryza sativa chromosome 10 BAC OSUNBao005K07 genomic sequence.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC087192; AAK84453.1; -.
DR Gramene; Q94GX7; -.
KW Hypothetical protein.
SQ SEQUENCE 184 AA; 19748 MW; B4E1FB4B28B050E CRC64;

Query Match
Best Local Similarity 1.0%; Score 8; DB 10; Length 184;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AAPAAPAK 148
DB 29 AAPAAPAK 36
|||||

RESULT 38
Q98BN4 PRELIMINARY; PRT; 215 AA.
ID Q98BN4; AC Q98BN4;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Cytochrome c.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Igesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003006; BAB51938.1; -.
DR InterPro; IPR000345; Cyt_c_heme_bind.
DR InterPro; IPR003088; Cyt_C1.
DR InterPro; IPR002327; Cyt_C1AB.
DR Pfam; PF00034; cytochrome_c; 1.
DR PRINTS; PR00604; CYTCHRMCIAB.
DR ProDom; PD000375; Cyt_C1AB; 1.

```



```

OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58."; 287:1809-1815(2000).
RL Science 287:1809-1815(2000).
DR EMBL; AE002534; AAF42176.1; -.
DR TIGR; NMB1841; -.
DR InterPro; IPR005835; NTP transferase.
DR Pfam; PF00483; NTP transferase; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 231 AA; 24537 MW; D02A35622345A189 CRC64;

```

```

Query Match      1.0%; Score 8; DB 16; Length 231;
Best Local Similarity 100.0%; Pred.No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      16 TAAFTQAS 23
      |||||
DB     116 TAAFTQAS 123

```

```

Search completed: December 9, 2003, 10:35:44
Job time : 69 secs

```